Top Curr Chem (2005) 247: 1–6
DOI 10.1007/b103817
© Springer-Verlag Berlin Heidelberg 2005

Collagens at a Glance

Jürgen Brinckmann (✉)

Department for Dermatology, University Lübeck, Ratzeburger Allee 160,
23538 Lübeck, Germany
brinckmann@molbiol.uni-luebeck.de

Abstract Collagens are a still growing family of proteins with an extraordinary heterogeneity: Today, 27 collagen types are known which are encoded by over 40 different genes. The diversity not only concerns the molecular assembly and the supramolecular structures but is also mirrored by tissue distribution, function and pathology of the collagen types. To become familiar with the collagen superfamily and to obtain a quick overview, we compiled Table 1 covering the "essentials" of the individual collagens.

1 Introduction 1

References 6

1 Introduction

Collagens are probably the most abundant proteins in the vertebrate body [1–3]. Their molecular hallmarks are the multiple repetitions of Gly-X-Y sequences and the unique triple helical structure built by three polypeptide chains. Up to now, 42 different polypeptide chains have been identified, which are encoded by 41 specific genes and compose 27 unique collagen types. The collagen family can be classified into different subfamilies according to their supramolecular assembly. The tissue distribution of the different collagen types discloses a remarkable diversity and range, for instance from a more exclusive pattern for collagen X in hypertrophic cartilage or collagen II in cartilage and vitreous to a ubiquitous occurrence of other fibrillar collagens such as collagens I and V. Information on the function of various collagens in a given tissue has been retrieved from studies of the macromolecular organisation of mutated collagens in heritable collagen diseases. Table 1 provides backbone information on chain composition, molecular and supramolecular assembly, as well as tissue distribution and pathology of the currently known collagen types.

Table 1

Type	Chain composition	Molecules	Subfamily	Tissue distribution (selection)	Pathology
I	$\alpha1$(I), $\alpha2$(I)	$[\alpha1(I)]_2\alpha2(I)$; $[\alpha1(I)]_3$	Fibrillar	Widespread: skin, bone, tendon, ligament, cornea	Osteogenesis imperfecta I–IV, Arthrochalatic type of Ehlers-Danlos syndrome (VIIA, VIIB), Classical type of Ehlers-Danlos syndrome
II	$\alpha1$(II)	$[\alpha1(II)]_3$	Fibrillar	Cartilage, vitreous	Achondrogenesis II, Hypochondrogenesis, Spondyloepiphyseal dysplasia congenita, Kniest dysplasia, Late onset Spondyloepiphyseal dysplasia, Stickler dysplasia
III	$\alpha1$(III)	$[\alpha1(III)]_3$	Fibrillar	Skin, vessel, intestine, uterus	Vascular type of Ehlers-Danlos syndrome, Hypermobile type of Ehlers-Danlos syndrome (rare), familiar aortic aneurysm
IV	$\alpha1$(IV), $\alpha2$(IV), $\alpha3$(IV), $\alpha4$(IV), $\alpha5$(IV), $\alpha6$(IV)	$[\alpha1(IV)]_2\alpha2(IV)$; $\alpha3(IV)\alpha4(IV)\alpha5(IV)$; $[\alpha5(IV)]_2\alpha6(IV)$	Network	Basement membranes	Alport syndrome, Benign familial hematuria
V	$\alpha1$(V), $\alpha2$(V), $\alpha3$(V), $\alpha4$(V) (rat)	$[\alpha1(V)]_2\alpha2(V)$; $[\alpha1(V)]_3$; $\alpha1(V)\alpha2(V)\alpha3(V)$	Fibrillar	Widespread: Bone, skin, cornea, placenta, Schwann cell (rat)	Classical type of Ehlers-Danlos syndrome

Table 1 (continued)

Type	Chain composition	Molecules	Subfamily	Tissue distribution (selection)	Pathology
VI	$\alpha1$(VI), $\alpha2$(VI), $\alpha3$(VI)	$\alpha1$(VI)$\alpha2$(VI)$\alpha3$(VI)	Network (also: microfibrils, broad banded structures)	Widespread: Bone, cartilage, cornea, skin, vessel	Bethlem myopathy, Ullrich congenital muscular dystrophy
VII	$\alpha1$(VII)	$[\alpha1(\text{VII})]_3$	Anchoring fibrils	Skin, bladder, oral mucosa, umbilical cord, amnion	Epidermolysis bullosa dystrophica
VIII	$\alpha1$(VIII), $\alpha2$(VIII)	$[\alpha1(\text{VIII})]_2\alpha2(\text{VIII})$ $[\alpha1(\text{VIII})]_3$ $[\alpha2(\text{VIII})]_3$	Network	Widespread: Descemet's membrane, vessel, bone, brain, heart, kidney, skin, cartilage	Fuchs endothelial corneal dystrophy
IX	$\alpha1$(IX), $\alpha2$(IX), $\alpha3$(IX)	$\alpha1$(IX)$\alpha2$(IX)$\alpha3$(IX)	FACIT*	Cartilage, cornea, vitreous	Multiple epiphyseal dysplasia
X	$\alpha1$(X)	$[\alpha1(\text{X})]_3$	Network	Hypertrophic cartilage	Schmid metaphyseal chondrodysplasia
XI	$\alpha1$(XI), $\alpha2$(XI), $\alpha3$(XI)	$\alpha1$(XI)$\alpha2$(XI)$\alpha3$(XI)	Fibrillar	Cartilage, intervertebral disc	Stickler syndrome type II, Marshall syndrome
XII	$\alpha1$(XII)	$[\alpha1(\text{XII})]_3$	FACIT*	Skin, tendon, cartilage	?
XIII	$\alpha1$(XIII)	Homotrimer ?	Transmembrane	Endothelial cells, skin, eye, heart, skeletal muscle	?
XIV	$\alpha1$(XIV)	$[\alpha1(\text{XIV})]_3$	FACIT*	Widespread: vessel, bone, skin, cartilage, eye, nerve, tendon, uterus	?

Table 1 (continued)

Type	Chain composition	Molecules	Subfamily	Tissue distribution (selection)	Pathology
XV	α1(XV)	Homotrimer ?	MULTIPLEXIN**	Capillaries, skin placenta, kidney, heart, ovary, testis	?
XVI	α1(XVI)	Homotrimer ?	FACIT*	Heart, kidney, smooth muscle, skin	?
XVII	α1(XVII)	$[\alpha 1(XVII)]_3$	Transmembrane	Hemidesmosomes in epithelia	Generalized atrophic benign epidermolysis bullosa, junctional epidermolysis bullosa localisata
XVIII	α1(XVIII)	Homotrimer ?	MULTIPLEXIN**	Perivascular basement membrane, kidney liver, lung	Knobloch syndrome
XIX	α1(XIX)	Homotrimer ?	FACIT*	Basement membrane zone in skeletal muscle, spleen, prostate, kidney, liver, placenta, colon, skin	?
XX	α1(XX), (chick)	Homotrimer ?	FACIT*	Corneal epithelium (chick)	?
XXI	α1(XXI)	Homotrimer ?	FACIT*	Vessel, heart, stomach, kidney, skeletal muscle, placenta	?

Table 1 (continued)

Type	Chain composition	Molecules	Subfamily	Tissue distribution (selection)	Pathology
XXII	α1(XXII)	Homotrimer ?	FACIT*	Tissue junctions (myotendinous junction in skeletal and heart muscle, articular cartilage-synovial fluid junction, border between the anagen hair follicle and the dermis in the skin	?
XXIII	α1(XXIII)	Homotrimer ?	Transmembrane	Heart, retina, metastatic cancer cells	?
XXIV	α1(XXIV)	Homotrimer ?	Fibrillar	Bone, cornea	?
XXV (CLAC-P***)	α1(XXV)	Homotrimer ?	Transmembrane	Brain, heart, testis, eye	?
XXVI	α1(XXVI)	Homotrimer ?	?	Testis, ovary	?
XXVII	α1(XXVII)	Homotrimer ?	Fibrillar	Cartilage	?

* Fibril associated collagens with interrupted tripelhelices.
** Multiple triple helix and interruptions.
*** Collagen-like Alzheimer amyloid component precursor.

References

1. Kadler K (1995) Protein Profile, Extracellular Matrix 1: Fibril-forming collagens
2. Kielty CM, Grant ME (2002) The collagen family: structure, assembly, and organization in the extracellular matrix. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders. Wiley-Liss, New York, p159
3. Ricard-Blum S, Dublet B, Rest M van der (2000) Protein profile, unconventional collagens: types VI, VII, VIII, IX, X, XII, XIV, XVI, and XIX. Oxford University Press

Top Curr Chem (2005) 247: 7–33
DOI 10.1007/b103818
© Springer-Verlag Berlin Heidelberg 2005

Structure, Stability and Folding of the Collagen Triple Helix

Jürgen Engel[1] (✉) · Hans Peter Bächinger[2]

[1] Department of Biophysical Chemistry, Biozentrum, University of Basel, 4056 Basel, Switzerland
juergen.engel@unibas.ch
[2] Shriners Hospital for Children, Research Department and Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, Oregon, 97239 USA

1 Structure of the Collagen Triple Helix 8

2 Collagens 12

3 Other Proteins with Collagen Triple Helices 14

4 Domains Involved in Chain Alignment 15

5 Model Peptides 16

6 The Coil to Triple Helix Transition 17

7 Thermodynamic Values for Stability 19

8 Mechanism of Stabilization by Proline and Hydroxyproline 23

9 *Cis-trans* Equilibria of Peptide Bonds 24

10 Kinetics of Triple Helix Formation 25

11 Hysteresis 28

12 Mutations in Collagen Triple Helices Effect Proper Folding 30

References 30

Abstract The collagen triple helix is a widespread structural element, which not only occurs in collagens but also in many other proteins. The triple helix consists of three identical or different polypeptide chains with the absolute requirement of a -Gly-Xaa-Yaa- repeat, in which the amino acid residues in X- and Y-position are frequently proline or hydroxyproline. The freezing of φ-angles in polypeptide backbone by proline rings and other steric restrictions are essential for stabilization. The OH-group of 4(*R*)-hydroxyproline, normally located in the Y-position, has an additional stabilizing effect. On the other hand, peptide bonds preceding proline and hydroxyproline are up to 20% in *cis*-configuration in unfolded chains and the need for a relatively slow *cis-trans* isomerization provides kinetic difficulties in triple helix formation. Because of their repeating structure, collagen chains easily misalign

during folding. Therefore, oligomerization domains flank triple helical domains in natural collagens. The mechanism by which these domains influence stability and kinetics was elucidated with model peptides using different types of trimerization domains. Finally, the review briefly describes mutations in collagen triple helices, which cause severe inherited diseases by disturbances of folding.

Keywords Collagen · Folding · Thermodynamics · Kinetics · Nucleation · Alignment · Oligomerization

List of Abbreviations

Ac	Acetyl
Ala (A)	L-Alanine
Arg (R)	L-Arginine
Asn (N)	L-Asparagine
Asp (D)	L-Aspartic acid
CD	Circular dichroism
Cys (C)	L-Cysteine
Flp	4(*R*)-Fluoroproline
Gln (Q)	L-Glutamine
Glu (E)	L-Glutamic acid
Gly (G)	Glycine
His (H)	L-Histidine
Hyp (O)	4(*R*)-L-Hydroxyproline
Ilu (I)	L-Isoleucine
Leu (L)	L-Leucine
Lys (K)	L-Lysine
Met (M)	L-Methionine
OMe	Methyl ester
Phe (F)	L-Phenylalanine
Pro (P)	L-Proline
Ser (S)	L-Serine
Thr (T)	L-Threonine
Tm	Midpoint transition temperature
Trp (W)	L-Tryptophane
Tyr (Y)	L-Tyrosine
Val (V)	L-Valine
Xaa, Yaa	Amino acid residue in X-, Y-position
ΔG^0	Standard free enthalpy (Gibbs free energy)
ΔH^0_{cal}	Standard calorimetric enthalpy
ΔH^0_{vH}	Standard van't Hoff enthalpy
ΔS^0	Standard entropy

1 Structure of the Collagen Triple Helix

The basic three-dimensional structure of the collagen triple-helix was first derived from fiber diffraction studies on collagen in tendon [1–3]. Each of the three chains in the molecule forms a left-handed polyproline-II-type helix. The

name of this helix is derived from poly-L-proline with all peptide bonds in *trans* conformation. It may however be formed also by peptides not containing proline or hydroxyproline. Its mode of stabilization is discussed later. The three helices, staggered by one residue relative to each other, are supercoiled along a common axis to a right-handed triple helix (Fig. 1).

A requirement for the formation of the superhelix is the repeating sequence -Gly-Xaa-Yaa-. A straight not supercoiled polyproline-II-helix has exactly three residues per turn and glycine residues are facing each other in the interior of the triple helix (Fig. 1B). Larger side chains at the C_α than H would prevent

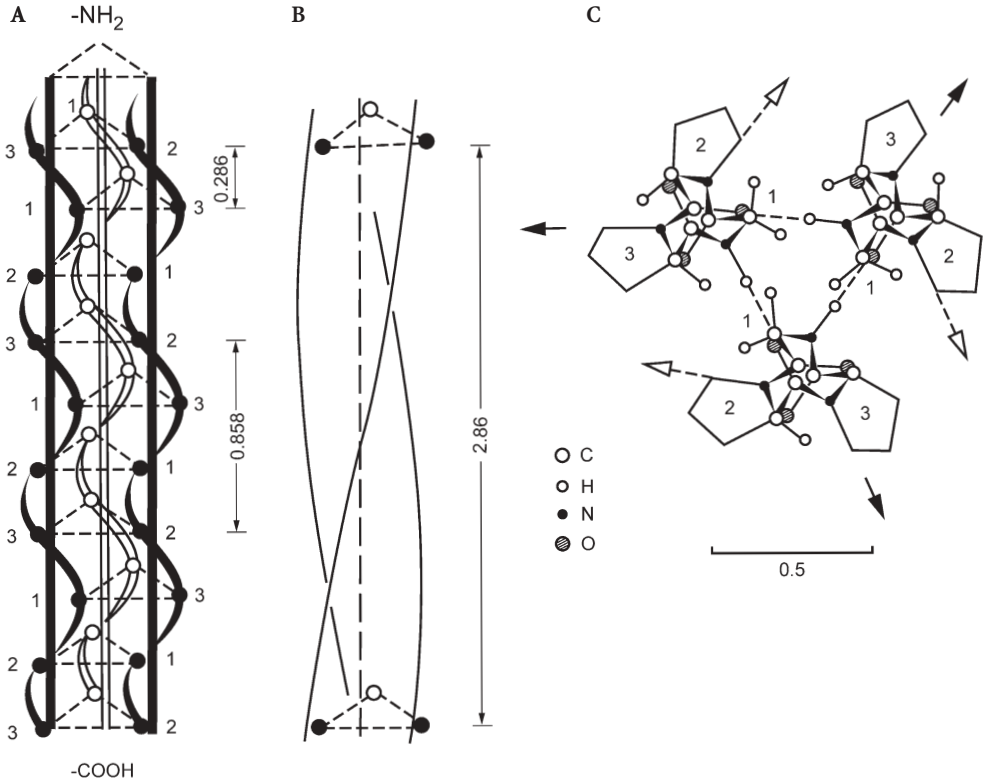

Fig. 1A–C Model of the collagen triple helix. The structure is shown for $(Gly\text{-}Pro\text{-}Pro)_n$ in which glycine is designated by 1, proline in X-position by 2 and proline in Y-position by 3: **A**, **B** side views. Three left-handed polyproline-II-type helices are arranged in parallel. For clarity the right-handed supercoil of the triple helix is not shown in A but indicated in B. Dashed lines indicate positions of C_α-atoms (and not hydrogen bonds as in C). All indicated values for axial repeats correspond to the supercoiled situation; **C** top view in the direction of the helix axis. The three chains are connected by hydrogen bonds between the backbone NH of glycine and the backbone CO of proline in Y-position (*dashed lines*). *Arrows* indicate the directions in which other side chains than proline rings emerge from the helix. Approximate residue-to-residue distances, repeats of the polyproline-II- and triple helix and a *scale bar* are indicated in nm

formation of a hydrogen bond between the backbone NH-group of glycine and the backbone CO of a residue in X-position of a neighboring chain. These hydrogen bonds are a major source of stability (see below). Isolated polyproline-II-helices are not stable if the polypeptide chains also contain other residues than proline and hydroxyproline. In contrast to the right-handed α-helices left-handed polyproline-II-helices are relatively rare structural elements in proteins with the striking exception of collagens.

The fiber diffraction, which were mainly performed on tendon fibers, showed a clear right-handed supercoiling of the three individual chains to a triple helix, leading to an increase to 3.33 residues per turn and a reduction of the axial repeat to 0.286 nm per residue, when viewed at the left handed individual helices (Fig. 1).

In each of the left handed helices, ten tripeptide units form three turns, and these helices were therefore called 10/3 helices. In the right handed triple helix the axial repeat is 2.86 nm because an identical structural element reoccurs after ten residues (Fig. 1B). Note that this element is located on a different chain. Lateral association of triple helices is very important in fibril formation and it should be noted already at this point that the interaction edges critically depend on the sequence of all three chains and on the symmetry of the helices. The collagen triple helix differs from other multi-stranded helices like coiled-coils structures or DNA double helices by a staggered arrangement of the three chains. As can be seen from Fig. 1A a glycine residue faces a residue in position X of a neighboring chain and this residue faces a residue in position Y of the third chain. Because of the stagger of the chains in collagen, for three different chains six alignments are possible. The topoisomerism of the triple helices is unknown for most natural collagens. It is however of importance for binding processes, in which an array of residues in different chains is recognized [4]. The staggered arrangement also leads to a bend between a collagen triple helix and a structure with a planar arrangement of chain ends. This bend was observed between $(GlyProPro)_{10}$ and foldon [5] and is predicted for scavenger receptor and Marco (see below).

Structural data were supplemented by crystallography of model peptides with Gly-Pro-Pro- and related repeats [6–9]. The earliest structure of the synthetic model peptide $(Pro\text{-}Pro\text{-}Gly)_{10}$ showed a 7/2 symmetry, giving rise to 3.5 residues per turn and an axial repeat of 0.2 nm [6]. This is in contrast to the result of fiber diffraction studies, which as mentioned showed a 10/3 symmetry. In 1994, the crystal of a peptide that consists of ten repeats of Pro-Hyp-Gly with a single substitution of a glycine residue by an alanine in the middle was analyzed [7]. This peptide was a triple helical molecule with a length of 8.7 nm and a diameter of about 1 nm. The structure with a resolution of 0.19 nm was consistent with the general parameters derived form the fiber diffraction model, but again showed a 7/2 symmetry. It was hypothesized that imino acid rich peptides have a 7/2 symmetry, but imino acid poor regions adopt a structure closer to a 10/3 helix [7]. This was later confirmed with a synthetic peptide that contained an imino acid poor region [9, 10]. The structure also confirmed

the expected hydrogen bond between the GlyNH and C=O of the proline in the Xaa position of a neighboring chain. In addition, an extensive water network was found, which formed hydrogen bonds with several carbonyl and hydroxyl groups of the peptide [11, 12]. This was interpreted as evidence for a stabilizing role of water, an issue that is highly controversial (see later).

Additional structural work was performed with peptides $(Pro\text{-}Pro\text{-}Gly)_{10}$ [13, 14], $(Pro\text{-}Hyp\text{-}Gly)_{10}$ [8] and $(Pro\text{-}Pro\text{-}Gly)_9$ [15]. In the later case the unusually high resolution of 0.1 nm was achieved. A structure of $(Pro\text{-}Pro\text{-}Gly)_{10}$ obtained from crystals grown in microgravity at the unusually high resolution of 0.13 nm showed a preferential distribution of proline backbone and side chain conformations, depending on the position [15–17]. Proline residues in the Xaa position exhibit an average main chain torsion angle of $\phi = -75°$ and a positive side-chain angle (down puckering of the proline ring) while proline residues in the Yaa position were characterized by a significantly smaller main chain torsion angle of –60° and a negative side chain angle (up puckering). These results were used to explain the stabilizing effect of hydroxyproline in the Yaa position, because hydroxyproline has a strong preference for up puckering. It also would explain the destabilizing effect of 4(*R*)-hydroxyproline in the Xaa position (see later).

A functionally important structural feature of the collagen triple helix is the orientation of side chains. Side chains of residues in X- and Y-positions point out of the helix and are freely accessible for binding interactions. In fibril formation, intermolecular interactions between oppositely charged residues are important and hydrophobic interactions between residues of different molecules, also play a role [19, 20]. The interactions between the collagen molecules in a fiber cause the characteristic quarter repeat of 67 nm and the complex cross-striation banding, both seen in electron micrographs [21, 22]. For quantification, the accurate distribution of residues at the interaction boundaries must be known. As mentioned, they are strongly dependent on helix symmetry. For simplification constant values of the pitch were assumed but different axial repeats ranging from 0.2 to 0.286 nm were used by different authors in such calculations. Not surprisingly very different interaction edges were predicted. In reality, the axial repeat may be rather variable depending on the sequence in a given region. As already mentioned, sequence dependent structural variations were indeed observed by crystallography [10]. In addition, the relatively strong intermolecular interactions may also alter the pitch and other structural parameters of the triple helix [23].

Several NMR-studies deal with the dynamics of the collagen triple [24–26] but many questions concerning the flexibility and fluctuations of the structure are still open. Several structures of synthetic model peptides were elucidated by NMR [25, 27–31]. Isotopic labeling was used to observe specific residues using heteronuclear NMR techniques. Hydrogen exchange studies were used to show that the NH groups of glycine exchanged faster in the imino acid poor region of a synthetic peptide compared to the Gly-Pro-Hyp region [25]. The hydration of the triple helix was also studied by NMR [32]. The hydration shell

was found to be kinetically labile with upper limits for water molecule residence times in the nanosecond to sub-nanosecond range.

The flexibility of the triple helix is apparent from electron micrographs of single molecules. Curved rods with a persistence length comparable to that of actin filaments were observed [33] and sites of increased bending were observable in wild-type [33] and mutated [34] molecules. Interestingly fiber forming collagen I is a rather straight rod, suitable for parallel lateral alignment in fibers. In contrast, the triple helix of the basement membrane collagen IV contains many bends, which are caused by interruptions in the regular Gly-Xaa-Yaa-repeat [35]. The increased flexibility is of functional relevance for the assembly of collagen IV to sheath-like structures.

2 Collagens

Collagens are very abundant proteins in the animal kingdom and are mainly located in the extracellular matrix. Typical locations are the mesogloea of jellyfish, the cuticle of worms, basement membranes of flies and the connective tissue of mammalians. Many of the collagens are well conserved from invertebrates to vertebrates, an example is the basement membrane collagen IV. There are also species specific collagens and the cuticle collagens may serve as examples.

The human collagen family of proteins now consists of 27 types. The individual members are numbered with roman numerals. The family is subdivided into different classes: the fibrillar collagens (types I*, II, III, V*, XI*, XXIV, and XXVII), basement membrane collagens (type IV*), fibril-associated collagens with interrupted triple helices (FACIT collagens, types IX*, XII, XIV, XVI, XIX, XX and XXI), short chain collagens (types VIII* and X), anchoring fibril collagen (type VII), multiplexins (types XV and XVIII), membrane associated collagens with interrupted triple helices (MACIT collagens, types XIII, XVII, XXIII, and XXV), and collagen type VI*. The types indicated by an asterisk are hetero trimers, consisting of two or three different polypeptide chains. Type IV collagens contain six different polypeptide chains that form at least three distinct molecules and type V collagens contain three polypeptide chains in probably three molecules.

The common feature of all collagens is the triple helical domain with the repeated sequence -Gly-Xaa-Yaa- in the primary structure and the high content of proline and hydroxyproline residues, respectively. A large variation in the length of the triple helix can be found in nature [36], the shortest collagen described is 14 nm long (minicollagen of the nematocyst wall in hydra) and the longest ones 2400 nm (cuticle collagen of annelids). Collagens however also contain variable numbers of non-collagenous domains (frequently abbreviated NC-domains) including globular domains (examples: Van Willebrand type A domains, fibronectin type III domains) and three-stranded coiled-coil regions. In their multidomain nature collagens do not differ from other multifunctional

modular extracellular matrix proteins. However, the elongated shape of the triple helical domains and the linear arrangement of solvent exposed sidechains (see previous section) lend special properties to collagens. They are well suited to form long fibrils or assemble to complex supramolecular structures in skin, bone, tendon, cartilage, basement membranes and other connective tissues. The types of supramolecular structures differ considerably for different collagen types and reviews can be found in [20, 37]. Besides their structural roles, collagens interact with numerous other molecules and are crucial for development and homeostasis of connective tissue. Much studied are the binding sites for numerous integrins, which are ubiquitous cellular receptors for extracellular matrix proteins.

The domain organization of a typical fibril forming collagen (collagen III) is shown in Fig. 2. This collagen consists of three identical chains. After removal

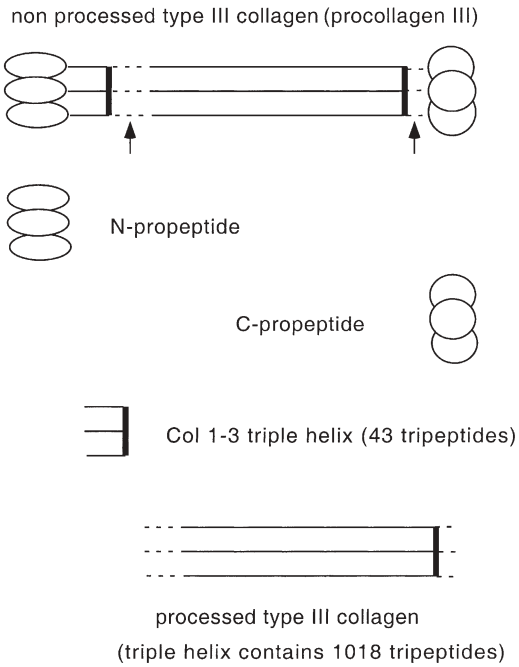

Fig. 2 Domain organization of collagen III. *At the top of the figure* collagen III is shown before processing by N- and C-proteinase, which cleave at sites marked by arrows after secretion. Triple helix forming regions with (GlyXaaYaa)-repeats are indicated by *solid vertical lines* and disulfide knots connecting the three identical chains by *vertical bars*. *Dashed lines* represent telopeptide regions with no (GlyXaaYaa)-repeats. Together with the N-propeptide the short triple helical region Col 1–3 is cleaved off, which served as model for triple helix folding. Many other folding studies were performed with mature processed collagen III (*lower part of the figure*), which contains 1018 tripeptide units in a triple helix. The number of tripeptide units in the triple helix is calculated from the number of tripeptides in each chain n by 3n–2, because 2 residue units are not incorporated. The chains are staggered by one residue per chain

of the signal peptide (not shown) it starts with an N-terminal non-collagenous domain also called N-propeptide, which is followed by a short triple helix with 10 Gly-Xaa-Yaa-tripetide units per chain. A short sequence of residues without a Gly-Xaa-Yaa-repeat links it to the central triple helix of 343 tripeptide units per chain. Both triple helices are terminated by disulfide knots, which interlink all three chains. The sequences involved are GSOGPOGICESCPT in the short and GPOGAOGPCCGG in the central triple helix. The disulfide knot is followed by a large C-terminal domain (C-propeptide), which is essential for chain registration and triple helix nucleation (see later).

3 Other Proteins with Collagen Triple Helices

Gly-Xaa-Yaa repeats that form triple helices are found in a number of proteins, which are not classified as collagens: complement protein C1q, lung surfactant proteins A and D, mannose binding protein, macrophage scavenger receptors A (types I, II, and III), MARCO, ectodysplasin-A, scavenger receptor with C-type lectin (SRCL), the ficolins (L, M and H), the asymmetric form of acetylcholinesterase, adiponectin, and hibernation proteins HP-20, 25, and 27. For these proteins, triple helices are usually much shorter than for collagens and the potential for fibril formation is less prominent. Instead, the triple helical domains may serve as spacer elements between other domains or as oligomerization domains, which bundle three or more functional domains to multivalent complexes. The triple helices may also contain binding sites for interactions with other binding partners as it was demonstrated for macrophage scavenger receptors.

C1q, a subunit of the first component of complement C1 [38] is a classical example for a functionally important oligomerization of binding domains by triple helices. The globular heads of C1q, which bind to immunoglobulins are composed of three domains connected by a collagen triple helix. Six trimers are linked by a collagen microfibril, leading to flower bouquet-like shape of the entire C1q molecule. The heads have a weak potential to bind to the Fc domains of IgG and IgM but binding of individual heads is too weak to mediate efficient interactions. Only when the heads are connected does C1q strongly interact with clusters of IgG. It is believed that the oligomeric structure of C1q is designed for efficient binding of clusters of IgG at an immunologically marked cell surface avoiding binding to isolated IgG, which would cause unwanted reactions with the complement system [39]. A similar effect is observed for other proteins in which oligomers are needed for high affinity [38]. It is well known that most lectins recognize monomeric sugars with only weak and polymeric structures with high affinity. In many cases, this physiologically important feature is generated by oligomerization of lectin domains. In an important class of lectins, the collectins oligomerization is achieved by collagen triple helices, which trimerize the C-type lectin domains at their C-termini [40, 41].

The distance between the carbohydrate binding sites is optimal for recognition of repeating units in a sugar chain or for cross-linking between chains [39, 42].

In macrophage scavenger receptor [43] the globular heads are connected to a collagen triple helix, which is followed by a three-stranded coiled coil. The two three-stranded structures probably stabilize each other in a mutual manner.

4
Domains Involved in Chain Alignment

Before turning to triple helix folding it is important to familiarize with domains flanking the triple helical regions. These domains are frequently called N- and C-terminal propeptides or more globally non-collagenous (NC-) domains. As an example see the domain organization of procollagen III in Fig. 2. Up to about 1980 it was believed that the few collagens known at this time (mainly collagen I) consist of a triple helix and short terminal so-called telopeptides only. Today it is known that the N- and C-terminal propeptides are still present at the time at which the three polypeptide chains assemble and fold to native collagen molecules. It will be shown in the following sections that they play an important role in chain alignment, selection of proper chain composition and nucleation of triple helix formation. It is therefore not surprising that experiments prior to the discovery of the propeptides on folding of collagen triple helices yielded data, which do not reflect the physiological situation. Early experiments suffered from poor refolding yields, misalignments of chains and formation of wrong products. A word of warning should be placed because even to-day commercially produced collagens without their natural propeptides are used for physical and chemical model studies without a clear recognition that important helper-domains are missing.

In the biosynthesis of fibrillar collagens the propeptides are proteolytically removed only briefly before fibril formation and these events plays an important regulatory role in this process. The liberated propeptides are stable structures and can be monitored for long periods of time in blood circulation and tissue distributions. Biological functions have been assigned to the circulating prodomains and deviations. In other collagens (collagens IV and VI) the NC-domains stay as part of the structure even after assembly.

In collagen III the amino-terminal end of the carboxy-terminal propeptide contains short coiled-coil segments with three to four heptad repeats that might facilitate the trimerization potential of this domain and the nucleation of the triple helix [44]. Coiled-coil structures are very frequently occurring oligomerization domains [45, 46]. The telopeptides also contain a lysine residues for the formation of covalent crosslinks between molecules. The globular amino-terminal propeptide (Fig. 2) potentially interacts with growth factors and is probably not involved in triple helix formation. Coiled-coil domains are prevalent in non-fibrillar collagens as well. The transmembrane domain containing collagens XIII, XVII, XXII and XXV show coiled-coil domains at the amino-ter-

minal end of the extracellular domain, close to the start of the triple helix and are apparently involved in alignment and nucleation [47].

5 Model Peptides

Synthetic peptides containing Gly-Xaa-Yaa repeats have been extensively used to study the thermal stability and folding of the collagen triple helix. These peptides can be synthesized as either single chains or crosslinked peptides. Solid-phase methods allowed the synthesis of peptides with a defined chainlength. $(Pro\text{-}Pro\text{-}Gly)_n$ with n=5, 10, 15, and 20 [48], $(Pro\text{-}4(R)Hyp\text{-}Gly)_n$ with n=5–10 [49], and $(Pro\text{-}Pro\text{-}Gly)_n$ with n=10, 12, 14, and 15 [50] were synthesized. These peptides were widely distributed and studied in several research groups. Work was also performed with $(Gly\text{-}Pro\text{-}Gly)_n$ with n=1–8 [51] and $(Gly\text{-}Pro\text{-}Pro)_n$ with n=3–7 [52]. Sutoh and Noda introduced the concept of block copolymers, where two blocks of $(Pro\text{-}Pro\text{-}Gly)_n$ n=5, 6, or 7 were separated by a block of $(Ala\text{-}Pro\text{-}Gly)_m$ m=5, 3, and 1, respectively [50]. This concept was later extended to include all amino acids and called host-guest peptide system. The most studied host-guest system uses the sequence Ac-$(Gly\text{-}Pro\text{-}4(R)Hyp)_3$-Gly-Xaa-Yaa-$(Gly\text{-}Pro\text{-}4(R)Hyp)_4$-Gly-Gly-$NH_2$ [53], but other variations were also studied [54, 55].

Because of the concentration dependence and the problem of formation of misaligned structures from single chain peptides, covalently linked homotrimeric peptides were synthesized by various methods. Propane tricarboxylate tris(pentachlorophenyl) and *N*-tris(6-amino hexanoyl)-lysyl-lysine was used as covalent bridging molecules for the three chains [56]. Other covalent linkers include Kemp triacid [57], a modified di-lysine system [58] and tris(2-aminoethyl)amine with succinic acid spacers [59].

Peptides with covalent crosslinks at both ends were also synthesized [60]. Peptide amphiphiles [61–63], and the iron(II) complex of bipyridine [64] have also been used to study trimeric peptides.

The sequence found at the carboxy-terminal end of type III collagen Gly-Pro-Cys-Cys-Gly (see above) was introduced in synthetic peptides or recombinantly expressed model proteins and trimerization was achieved by reoxidation in the presence of mixtures of reduced and oxidized gluthatione [30, 31, 65–68]. The non-covalent obligatory trimer foldon [67, 69] was fused to the C-terminal and N-terminus of model peptides of recombinant model proteins and in this way a very effective trimerization was achieved. Very recently a larger version of foldon (minifibritin) was combined with a collagen III-like disulfide knot to facilitate the oxidative disulfide coupling [70]. Initially, all peptides were studied as homotrimers, but strategies were developed to synthesize heterotrimers. Fields proposed the use of the di-lysine system to synthesize heterotrimeric peptides of type IV collagen [58]. Moroder used a simplified cysteine bridge to synthesize the collagenase cleavage site of type I collagen [71] and the integrin binding site of type IV collagen [72].

6
The Coil to Triple Helix Transition

Model peptides, which form triple helices, have been instrumental for the elucidation of the mechanism of the coil ⇆ triple helix transition and stabilizing interactions because of their uniform sequence and short length. We shall therefore describe their transition prior to those of long intact collagens. For short chains and a high cooperativity intermediates in a sequential folding process, can be neglected and the transition may be approximated by an "all-or-none" process from randomly coiled species to a fully helical species H with a single equilibrium constant [73]:

$$K = \frac{[H]}{[C]^3} = \frac{F}{3c_0^2\,(1-F)^3} \tag{1}$$

Here c_0=3 [H]+[C] is the total concentration of chains and F=3[H]/c_0 the degree of helicity. Each triple helix H contains 3n–2 tripeptide units in helical state. Subtraction of 2 originates from the chain stagger by which 2 tripeptide units cannot fully participate in the triple helix. Concentrations of H and C may also be expressed in concentrations of tripeptide units in either helical or coiled state.

From

$$\Delta G^0 = -RT\ln K = \Delta H^0 - T\Delta S^0 \tag{2}$$

and from Eq. (1) it follows that at the midpoint of the transition (F=0.5 and T=T_m),

$$T_m = \frac{\Delta H^0}{\Delta S^0 + R\ln(0.75c_0^2)} \tag{3}$$

where ΔG^0 is the standard Gibbs free energy, ΔH^0 is the standard enthalpy, and ΔS^0 is the standard entropy. It is important to note that the T_m is concentration dependant and should only be compared at identical molar chain concentrations or concentrations which were corrected for concentration differences with Eq. (3).

ΔH^0 can be obtained from the slope of the transition curves. The van't Hoff enthalpy can be approximated from the slope of the transition curve at its midpoint:

$$\Delta H^0 = 8\,RT_m^2\left(\frac{dF}{dT}\right)_{F=0.5} \tag{4}$$

or more accurately by fitting of the entire transition curve with Eq. (1) and

$$K = \exp\left[\frac{\Delta H^0}{RT}\left(\frac{T}{T_m}-1\right) - \ln 0.75c_0^2\right] \tag{5}$$

This relation is obtained by solving Eq. (3) for ΔS^0 and substituting for ΔS^0 in Eq. (2). The temperature dependence of ΔH^0 is neglected, because the specific heat of the coil ⇆ triple helix transition was found to be close to 0 [74]. F is determined from the molar ellipticity [Θ] of CD or any other signal proportional to helicity. When measuring CD the molar ellipticity is given by

$$[\Theta] = F([\Theta]_h - [\Theta]_c) + [\Theta]_c$$

$[\Theta]_h$ and $[\Theta]_c$ are the ellipticities for the peptide in either the completely triple helical conformation or in the unfolded conformation. These values normally show linear temperature dependencies which can be described by

$$[\Theta]_h = [\Theta]_{h,Tm} + p(T - T_m)$$

$$[\Theta]_c = [\Theta]_{c,Tm} + q(T - T_m)$$

Here the ellipticities at the midpoint temperature T_m are arbitrarily used as reference points. For a best fit of the transition curves, initially the T_m is kept constant and ΔH^0, p and q are varied. After an initial fit, the T_m is then also varied.

For peptides that contain a cross-link, the triple helix ⇆ coil transition becomes concentration independent and Eqs. (1) and (3) simplify to

$$K = \frac{F}{1 - F}$$

and

$$T_m = \frac{\Delta H^0}{\Delta S^0}$$

The fitting procedure for the determination of the thermodynamic parameters is otherwise the same as for non-linked chains.

ΔH^0 can also be determined calorimetrically. Most peptides, for which calorimetric data is available, show that the ratio of the van't Hoff enthalpy and the calorimetric enthalpy is close to 1, indicating that the all-or-none mechanism is a good approximation. However, exceptions have been reported, especially for heterotrimeric peptides [72]. Great care has to be taken to evaluate true equilibrium transition curves by performing measurement at sufficiently slow speed. Because of the slow folding process in the transition region a hysteresis and slow establishment of equilibrium is observed, which is demonstrated for Ac-(Gly-Pro-4(*R*)Hyp)$_{10}$-NH_2 in Fig. 3. The heating and cooling rate was as low as 10 °C/h but large deviations between transition curves recorded by heating and cooling were observed. Differences are dramatic at low concentrations.

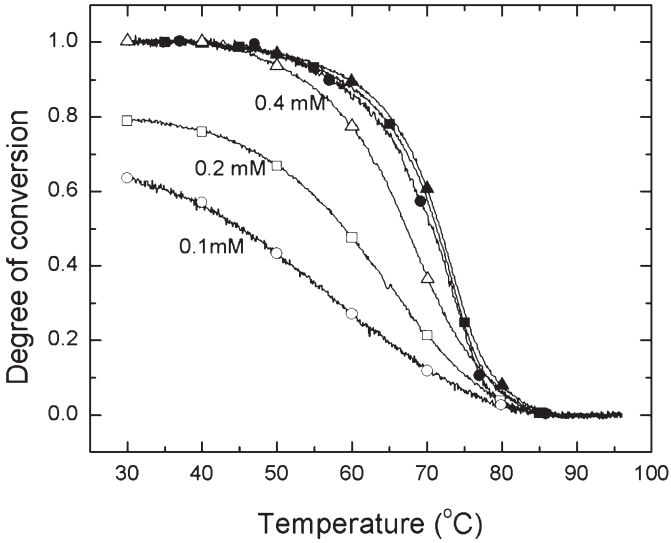

Fig. 3 Triple helix coil transition of Ac-(Gly-Pro-4(*R*)Hyp)$_{10}$-NH$_2$ in water. The transition was monitored by circular dichroism at 225 nm. *Closed symbols* are for heating and *open symbols* for cooling. The heating/cooling rate was 10 °C/h

7 Thermodynamic Values for Stability

The temperature at the midpoint of the transition T_m is frequently used as a measure of stability in comparisons of different peptides or collagens. As shown above this value is dependent on the concentration of the peptide and the rate of heating, and only values either measured or corrected to the same concentration and heating rate should be compared. Unfortunately in most publications concentrations are not included. Published thermodynamic data for the triple helix coil transitions of identical synthetic peptides vary significantly (Tables 1 and 2). The commercially available peptides (Pro-Pro-Gly)$_{10}$ and (Pro-4(*R*)Hyp-Gly)$_{10}$ have been measured in several laboratories. The van't Hoff enthalpy change varies from –7.91 to –18.8 kJ/mol tripeptide units for (Pro-Pro-Gly)$_{10}$ and from –13.4 to –25.25 kJ/mol tripeptide units for (Pro-4(*R*)Hyp-Gly)$_{10}$. These deviations point to the above mentioned problems of achieving equilibrium and to other sources of error including unknown concentration dependencies, misalignments of single chain peptides and missing checks of the validity of the "all-or-none" approximation. It is recommended to consult the individual publications in order to gain information on reliability.

Table 1 presents a selection of data on model peptides with repeating Gly-Xaa-Yaa units and Table 2 contains selected data on host-guest peptides. Note that the values of ΔH^0_{vH}, ΔH^0_{cal} and ΔS^0 of the homotripeptides in Table 1 are

Table 1 Thermodynamic data of the triple helix coil transition of model peptides in aqueous solution determined from transition curves ($\Delta H^0_{vH}, \Delta S^0$) and differential scanning calorimetry (ΔH^0_{cal})

Peptide	T_m (°C)	ΔH^0_{vH} (kJ/mol tripeptide)	ΔS^0 (J/mol tripeptide/K)	ΔH^0_{cal} (kJ/mol tripeptide)	Reference
$(Pro\text{-}Pro\text{-}Gly)_{10}$	28	−10.6	−26.5		[50]
$(Pro\text{-}Pro\text{-}Gly)_{10}$	24.6	−7.91	−22.4	−7.68	[73]
$(Pro\text{-}Pro\text{-}Gly)_{10}$	25	−18.8	−58.6		[75]
$(Pro\text{-}Pro\text{-}Gly)_{10}$	34	−18.6	−60.9		[76]
$(Pro\text{-}Pro\text{-}Gly)_n$ n=10, 15, 20		−8.2	−19.2		[77]
$(Pro\text{-}Pro\text{-}Gly)_n$ n=12, 14, 15		−10.6	−26.5		[50]
$(Pro\text{-}Pro\text{-}Gly)_{10}$	32.6			−6.43	[78]
$(Pro\text{-}4(R)Hyp\text{-}Gly)_{10}$	57.3	−13.4	−36.4	−13.4	[73]
$(Pro\text{-}4(R)Hyp\text{-}Gly)_{10}$, pH 1	60.8	−23.6	−65.6		[79]
$(Pro\text{-}4(R)Hyp\text{-}Gly)_{10}$, pH 7	57.8	−23.3	−65.2		[79]
$(Pro\text{-}4(R)Hyp\text{-}Gly)_{10}$, pH 13	60.8	−25.1	−70.7		[79]
$(Pro\text{-}4(R)Hyp\text{-}Gly)_{10}$	60.0			−13.9	[78]
$(Pro\text{-}4(R)Flp\text{-}Gly)_{10}$	80	−17.1			[76]
$Ac\text{-}(Gly\text{-}4(R)Hyp\text{-}Thr)_{10}\text{-}NH_2$	18	−27.1	−87.1	−27.5	[80]
$Ac\text{-}(Gly\text{-}Pro\text{-}Thr(\beta Gal)_{10}\text{-}NH_2$	38.8	−14.0	−39.6		[80]
$Ac\text{-}(Gly\text{-}4(R)Hyp\text{-}Thr(\beta Gal)_{10}\text{-}NH_2$	50.0	−11.1	−29.0		[80]
$((Gly\text{-}Pro\text{-}Thr)_{10}\text{-}Gly\text{-}Pro\text{-}Cys\text{-}Cys)_3$	13.8	−7.1	−41.5	−11.9	[81]
$((Gly\text{-}Pro\text{-}Pro)_{10}\text{-}Gly\text{-}Pro\text{-}Cys\text{-}Cys)_3$	82				[69]
$((Gly\text{-}Pro\text{-}Pro)_{10})_3$ foldon	66			−10.7	[67]
	70				[69]

expressed per mol tripeptide units in order to compare data of peptides with different length. It has to be remembered, however, that contributions of tripeptide units might not be completely additive because of energies, involved in nucleation of the triple helix [85]. For the host-guest peptides with different equences in the guest and host only total energies per mol triple helix are given in Table 2. Thermodynamic values are dominated by the large proline- and hydroxyproline-rich host and the influence of the guest tripeptide is rather small. The effects of tripeptides of interest are only noticed by comparing T_m and other thermodynamic values with those of a reference guest peptide. T_m values are normally used as a measure of stability, although in a strict thermodynamic sense standard values of Gibbs free energy ΔG^0 at 273.2 K (25 °C) should be used. Such values have not been evaluated for model peptides and only values at the midpoint temperature can be calculated by $\Delta G^0_{Tm}=\Delta H^0-T_m\Delta S^0$.

The data summarized in Tables 1 and 2 clearly confirm the crucial role of imino acids and in particularly of 4(*R*)Hyp in Y-position in the stabilization of the triple helix. The dominating effects of proline and hydroxyproline in the flanking Gly-Pro-4(*R*)Hyp sequences of the host/guest peptides facilitate triple helix formation even for highly destabilizing guests. Clearly tripeptide

Table 2 Thermodynamic data of the triple helix formation of host-guest peptides in PBS determined from transition curves (ΔH^0_{vH}, ΔS^0) and differential scanning calorimetry (ΔH^0_{cal})

Peptide	T_m (°C)	ΔH^0_{vH} (kJ/mol triple helix)	ΔS^0 (J/mol triple helix/K)	ΔH^0_{cal} (kJ/mol triple helix)	Reference
HG[a] Gly-Pro-4(*R*)Flp	43.7			–204	[78]
HG Gly-4(*R*)Hyp-Pro	43.0			–204	[78]
HG Gly-4(*R*)Hyp-4(*R*)Hyp	47.3			–217	[78]
HG Gly-Pro-Pro	45.5			–213	[78]
HG Gly-Ala-4(*R*)Hyp	39.9	–423.7	–1214		[53]
HG Gly-Ala-4(*R*)Hyp	41.7	–480			[82]
HG Gly-Leu-4(*R*)Hyp	39.0	–437.5	–1256		[53]
HG Gly-Phe-4(*R*)Hyp	33.5	–514.1	–1549		[53]
HG Gly-Pro-Ala	38.3	–358.0	–1005		[53]
HG Gly-Pro-Ala	40.9	–502			[82]
HG Gly-Pro-Leu	32.7	–514.6	–1549		[53]
HG Gly-Pro-Leu	31.7	–514			[82]
HG Gly-Pro-Phe	28.3	–557.3	–1717		[53]
HG Gly-Pro-Arg	47.2	–610	–1100		[83]
HG Gly-Pro-Arg pH 2.7	45.5	–560	–1300		[83]
HG Gly-Pro-Arg pH 12.2	43.1	–390	–1100		[83]
HG Gly-Asp-Lys pH 2.7	26.5	–600	–1000		[83]
HG Gly-Asp-Lys pH 12.2	29.9	–490	–1500		[83]
HG Gly-Asp-Arg pH 2.7	33.4	–550	–1700		[83]
HG Gly-Asp-Arg pH 12.2	34.4	–460	–1300		[83]
HG Gly-Glu-Lys pH 2.7	29.5	–490	–1500		[83]
HG Gly-Glu-Lys pH 12.2	33.1	–530	–1600		[83]
HG Gly-Glu-Arg pH 2.7	37.3	–530	–1600		[83]
HG Gly-Glu-Arg pH 12.2	39.1	–510	–1500		[83]
HG Gly-Lys-Asp pH 2.7	30.5	–770	–2400		[83]
HG Gly-Lys-Asp pH 12.2	30.2	–720	–2300		[83]
HG Gly-Arg-Asp pH 2.7	28.8	–720	–2300		[83]
HG Gly-Arg-Asp pH 12.2	31.9	–630	–1900		[83]
HG Gly-Lys-Glu pH 2.7	36.5	–620	–1900		[83]
HG Gly-Lys-Glu pH 12.2	31.6	–680	–2100		[83]
HG Gly-Arg-Glu pH 2.7	35.0	–630	–1900		[83]
HG Gly-Arg-Glu pH 12.2	32.2	–710	–2200		[83]
HG Gly-Gly-Phe	19.7	–647	–2093		[84]
HG Gly-Gly-Leu	23.9	–578	–1800		[84]
HG Gly-Gly-Ala	25.0	–559	–1758		[84]
HG Gly-Ala-Ala	29.3	–450	–1400		[83]
HG Gly-Ala-Leu	27.8	–574	–1758		[53]
HG Gly-Phe-Ala	23.4	–593	–1884		[53]
HG Gly-Ala-Phe	20.7	–637	–2051		[53]

[a] HG refers to the host-guest peptides with the structure Ac(Pro-4(*R*)Hyp-Gly)$_3$-Gly-Xaa-Yaa-(Pro-4(*R*)Hyp-Gly)$_4$-Gly-Gly-NH_2. The tripeptide sequence given after HG corresponds to Gly-Xaa-Yaa. The peptides were measured in PBS, pH 7, unless indicated otherwise.

units which lack Pro and 4(*R*)Hyp contribute little energy and homopeptides containing these units are very instable or do not form triple helices at all. Replacement of 4(*R*)Hyp in Y-position by Pro leads to a decrease of T_m and to decreases of standard values of negative enthalpy. The mode of stabilization by proline and hydroxyproline will be discussed in the next paragraph. An interesting additional stabilization was discovered in deep-sea annelid collagens [65] and followed up with model peptides by Bann et al. [80] (Table 1). *O*-Glycosylation of threonine in position Y stabilizes the triple helix in a similar way as hydroxyproline. Glycosylation of the OH-groups in the annelid collagen was demonstrated by amino acid sequencing but the nature of the sugars remained unknown. In these peptides one galactose was attached to threonine with a β-linkage. Interestingly unglycosylated threonine has no stabilizing effect. Another residue with high stabilization potential is arginine in position Y [83]. There are also data, which suggest that glutamine in position X and arginine in Y form stabilizing pairs.

More general attempts have been made to predict the stability contributions of different peptide units in collagens from thermodynamic data of peptides. The pioneering work of Dölz and Heidemann [86] and two recent papers may serve as examples [78, 87].

When comparing the values in Tables 1 and 2 in detail many conflicts can be discovered and the limitations discussed in the beginning of this section should be kept in mind. Only few thermodynamic data have been collected for trimeric peptides generated by fusion with the disulfide knot of collagen III or a trimeric registration domain. The very stable trimeric foldon domain of T4-phage fibritin was successfully applied (Table 1) [67, 69]. For trimerized model peptides the T_m is much higher than for single chain peptides and is concentration independent. The oligomerization and registration of the chains leads to a stabilization by entropic effects [67, 88]. At the site of linkage the three chains are held in very close neighborhood. The local concentration at this site was estimated to be close to 1 mol/l. The stabilization by foldon resembles the function of the oligomerization domains in natural collagens. Recently foldon was linked to human type I and III collagen and efficient trimerization was observed [89]. Unexpectedly a large increase in yields of expression from a yeast system was also achieved. Trimerized peptides offer large advantages and are much closer to real collagens than single chain peptides. In future comparisons of stability the risk of errors from unknown concentrations, misalignments and hysteresis effects could be avoided with these models.

Tables 1 and 2 only contain thermodynamic values of model peptides. Data also exist for collagens and fragments derived from them. These are averages of all contributions in a protein with a complicated sequence. It may be mentioned however, that the enthalpy change of the coil to triple helix transition of collagen I was determined to be ΔH^0=–(15–18) kJ/mol tripetide units [74]. It was also found that the specific heat of the reaction was near to zero. In contrast to globular proteins ΔH^0 is therefore temperature independent, a feature also found for some of the model peptides. Another general feature was noticed

by Privalov [74]. On a per residue basis, the change of enthalpy for structure formation of the triple helix is significantly larger than that of globular proteins.

8
Mechanism of Stabilization by Proline and Hydroxyproline

The two imino acids proline and hydroxyproline stabilize polyproline-II-type helices by specific steric restrictions imposed by the proline rings, which connect the Cα-atoms with the preceding nitrogen in the polypeptide backbone. The rotation around the Cα-N bond (Φ-angle) is thus frozen by the ring to about Φ=–75±35°. In addition rotation around the Cα-CO bond (Ψ-angle) is also heavily restricted by several clashes caused by unfavorable van der Waals interactions [90]. It was demonstrated that only Ψ=140° is accessible for poly-L-proline or poly-L-hydroxyproline with all peptide bonds in *trans* conformation. The two angles Φ and Ψ define the conformation of a polypeptide backbone uniquely. In a Ramachandran plot (Φ vs Ψ) [91, 92] the polyproline-II-helix exists only in a narrow area defined by these angles. Note that in the present review the newly defined zero origins for Φ and Ψ are used, whereas the excellent book by Dickerson and Geis still uses the old ones. The stability of the polyproline helix is largely defined by the intramolecular van der Waals interactions, which are responsible for the restrictions in conformational freedom. This fact is often overlooked in discussions about other modes of stabilization (hydrogen bonds, hydrophobic and electrostatic interactions). For iminoacids, however, the steric restrictions imposed by the proline ring play the dominant role.

Very early it was noticed that hydroxyproline exhibits an additional stabilizing effect. This followed from a comparison of the thermal stability of collagens from different species with different contents of proline and hydroxyproline [74, 93] and more conclusively from a comparison of collagen-like model peptides (see previous section). Only 4(*R*)-hydroxyproline [94], but not 4(*S*)-hydroxyproline [55, 95, 96] shows a stabilizing action indicating that the effect originates from the OH-group and its position at the proline ring.

Two very different explanations were offered to explain the stabilizing effect of the OH-group of 4(*R*)-hydroxyproline. The first explanation is a stabilizing effect of water bridges between the OH-group and backbone groups. Water mediated bridges were seen in crystal structures of model peptides ([11] and citations therein) but it remained open, whether these indirect hydrogen bond networks provide stabilizing energy to the peptide. Clearly large unfavorable entropic contributions are expected [97]. The observation that the stability difference between $(\text{Gly-Pro-Pro})_n$ and $(\text{Gly-Pro-Hyp})_n$ was maintained in non-aqueous solvents even after careful removal of trace water [85] also argued against the water bridge hypothesis. The second and now more generally accepted explanation is stabilization by the inductive effect of the OH-group.

This notion is mainly based on the finding that 4(*R*)-fluoroproline in the Y-position of the Gly-Pro-Yaa repeat, increases the midpoint transition temperature to even higher values than 4(*R*)-hydroxyproline [98]. The fluorine group does not form hydrogen bonds but has a higher electronegativity than the OH group.

The replacement of the hydroxyl group by fluorine has several effects. The inductive effect of the fluorine group influences the pucker of the pyrrolidine ring. This is a stereoelectronic effect as it depends on the configuration of the substituent. 4(*R*) Substituents stabilize the Cγ-exo pucker, while 4(*S*) substituents stabilize the Cγ-endo pucker. The X-ray structure of $(\text{Pro-Pro-Gly})_{10}$ showed a preference of proline residues in the Xaa position in Cγ-endo pucker, while the Yaa position prolines preferred Cγ-exo puckering [17, 18]. This puckering influences the range of the main chain dihedral angles φ and ψ of proline, which are required for optimal packing in the triple helix. In addition, the *trans*/*cis* ratio of the peptide bond is influenced by the substituent [99–101]. Because all peptide bonds in the triple helix have to be *trans*, the helix is stabilized, if the amount of *trans* peptide bonds in the unfolded state is increased.

9 *Cis-trans* Equilibria of Peptide Bonds

Before dealing with the kinetics of collagen folding in the following sections it is necessary to discuss the role of *cis-trans* isomerization of peptide bonds. The peptide bond shows partial double bond character and is planar [91]. The flanking C_α atoms can be either in *trans* (ω=180°) or *cis* (ω=0°) conformation. For peptide bonds preceding residues other than proline, only a small fraction (0.11 to 0.48%) was found to be in the *cis* conformation in the unfolded state [102]. For peptide bonds preceding proline this fraction is much higher (Table 3). The small energy difference between the two conformations is explained by similar internal van der Waals interaction in *cis* and *trans* conformation because of the symmetry of the tertiary peptide nitrogen. In contrast, for residues with side chains other than proline a large asymmetry exist between the N-C_α bond and the NH group. *cis* Contents from 10 to 30% were found in short unstructured proline containing peptides [103, 104]. This relates to the equilibrium constant between *cis* and *trans* states in the unfolded state $K_{cis/trans}$ of 0.11 to 0.43. Values of $K_{cis/trans}$ for a number of collagen models are listed in Table 3.

The collagen triple helix can accommodate only *trans* peptide bonds, so the ratio of *cis* to *trans* peptide bonds in the unfolded state has a direct influence on the stability of the triple helix.

In a transition from the unfolded state to the triple helix all *cis* peptide bonds have to be converted to *trans* because *cis* bonds can't be incorporated in this structure. Therefore, changes in $K_{cis/trans}$ may influence stability. Changes of the *cis-trans* equilibrium by inductive effects in hydroxyproline or fluoroproline (see above) are however relatively small. Much larger effects of *cis-trans* isomerization are expected for the kinetics of the transition, its high activation

Table 3 $K_{cis/trans}$ in model compounds and unfolded type I collagen measured by NMR

Compound	$K_{cis/transt}$	Reference
Ac-4(*S*) Hyp-OMe	0.37	Bächinger and Peyton
Ac-Pro-OMe	0.16	Unpublished
Ac-4(*R*)Hyp-OMe	0.12	
Ac-Pro-OMe	0.217	[100]
Ac-4(*R*)Hyp-OMe	0.164	
Ac-4(*R*)Flp-OMe	0.149	
Ac-4(*S*)Flp-OMe	0.4	
Ac-4(*S*)Hyp-OMe	0.417	[99]
Ac-Pro-OMe	0.21	[101]
Ac-4(*R*)Flp-OMe	0.14	
Ac-4(*S*)Flp-OMe	0.39	
Ac-4,4 difluoroproline-OMe	0.29	
Unfolded type I collagen		[105]
Xaa-Pro	0.19	
Xaa-Hyp	0.087	

energy (see following section). Kinetic parameters are also more significantly influenced by inductive effects of fluoroproline than equilibrium constants [101].

10 Kinetics of Triple Helix Formation

For a discussion of the kinetic mechanism of triple helix formation several steps have to be distinguished. Similar to the folding of other structures like α-helices, DNA-double helices or multi-stranded coiled-coil structures the nucleation of the helix is distinctly different from the following propagation steps. Nucleation includes the first encounter of chains, chain selection and the difficult and usually slow formation of a first nucleus of triple helical structure. Since the probability of meeting of three chains is very low in solution phase dimeric intermediates likely occur during nucleation. Nucleation may happen at many sites but may be limited to preferred sites of increased nucleation potential in other cases. Depending on whether nucleation or propagation is the rate limiting step, kinetics will be completely different. For the folding of the triple helix from single chains of unlinked model peptides or collagen fragments nucleation events are predominant at low peptide concentrations. For $(ProProGly)_{10}$ and $(ProHypGly)_{10}$ a reaction mechanism with a dimeric intermediate was derived [68] and the reaction order decreased from 3 (at very low concentrations) to 1 (extrapolated for high concentrations). The change of reaction order was explained by a switch of rate determining nucleation at low

Table 4 Rate constants and activation energies of triple helix folding from single and trimerized chains at 20 °C

Protein	k_a (M^{-2} s^{-1})	k (s^{-1})	E_a (kJ/mol)
$(ProProGly)_{10}$	900	–	7
$(ProHypGly)_{10}$	About 10^6	–	8
$(GlyProPro)_{10}$ foldon		0.00197	54.5
$(GlyProPro)_{10}Cys_2$		0.00033	52.5

to rate determining propagation at high concentrations. Rate constants of nucleation k_a and propagation k as well as activation energies E_a derived for these model peptides are listed in Table 4. Another kinetic study with a different single chain model was interpreted by a different mechanism of nucleation [106]. Both studies demonstrate the complexity of the kinetics of folding from single chains.

For this reason further kinetic work was performed with model peptides in which the three chains are linked, either by the disulfide knot of collagen III or by the obligatory trimer foldon [67–70]. For these models, which are much closer to true collagens, first order kinetics was observed and helix propagation was rate determining. Therefore, only propagation rate constants k could be determined (Table 4).

Most, if not all, collagens and collagen-like proteins contain a oligomerization domain or disulfide knot (see above), which keep the chains in an aligned trimeric state. These linkers provide the correct chain alignment and provide a high local concentration near the connecting point, thus facilitating nucleation. Nucleation is therefore much faster than propagation and is not observed in the kinetic process. In most collagens the oligomerization domains are located at the C-terminal end of the triple helical domain and propagation proceeds in a zipper-like fashion from the C- to N-terminus. This directionality of folding was very clearly demonstrated for collagen III by monitoring the growth of trypsin resistant helical segments [107]. Here folding starts at the disulfide knot (Fig. 2). Conclusive data were also obtained for collagen IV, in which propagation starts at the C-terminal propeptide [108]. In this case it was possible to visualize the growth of the triple helix directly by electron microscopy. The zipper like growth was confirmed by monitoring the chain length dependence of the folding kinetics [107, 109, 110]. The short and the major triple helical domains of collagen III were employed and supplemented by the three-quarter fragment of this molecule (Fig. 4).

Interestingly the kinetics proceeds in two phases. In the kinetic experiments of Fig. 4 the dead time was 25 s and therefore the fast phase remained unresolved. Its amplitude was about 50% of the total for the small triple helix of peptide Col 1–3 but was hardly detectable for the larger proteins. Here almost the entire transition proceeded in a slow phase. As expected for a zipper-like folding initial rates of the slow phase were inversely proportional to the number of

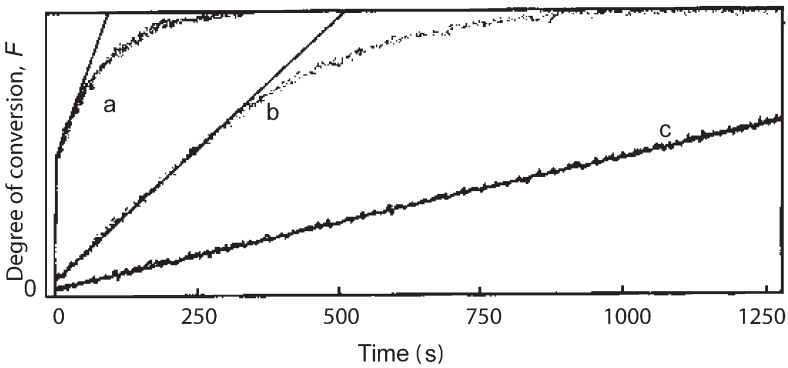

Fig. 4 Comparison of the folding kinetics of type III pN-collagen, the quarter fragment of type III collagen, and peptide Col 1–3. All three proteins contained a disulfide knot at the C-terminus. Refolding of the collagen (c), the quarter fragment (b), and Col1–3 (a) were measured by circular dichroism. The *straight lines* represent the initial rates. Original data from [110]

tripeptide units in the triple helices. Because of a long linear phase of the reaction [110] half times of folding increased proportionally with the number of tripeptide units.

The rate determining steps in the slow phase of propagation are *cis-trans* isomerization steps of peptide bonds. As shown in the previous section and Table 3, a large fraction of the peptide bonds preceding proline and hydroxyproline is in *cis*-conformation at equilibrium in the unfolded state. In collagen III such peptide bond alternate with those of much lower *cis* potential but at average the fraction of *cis*-bonds is very high. During folding *cis*-peptide bonds have to be converted to the trans-state required for the native triple helix. The rate constant of *cis* to *trans* conversion is approximately $k_{cis\text{-}trans}$= 0.0003 s^{-1} at 20 °C (h in Table 4) and similar rate constants were derived from an analysis of the kinetics of collagen [110].

A second characteristic feature of a kinetics determined by *cis-trans* isomerization is the high temperature dependence. For collagen III activation energies of E_a=44–70 kJ/mol were derived [107, 111]. The values match activation energies, which were determined for small proline dependent peptide [103, 104] and collagen model peptides [68]. In the transition from *cis* to *trans* the double bond related energy is removed at the transition state giving rise to a high activation energy. In a first approximation, the value of E_a is therefore an intrinsic parameter of the peptide bond but dependencies of side chains were noticed in studies with short peptides.

The fast phase of the folding process was correlated to the folding of the short region of the polypeptide chains before a peptide bond in *cis* conformation is met during zipper-like folding [110]. The amplitude was found to increase, when refolding was started from a non-equilibrium state of the unfolded molecules in which less *cis*-peptide bond were present than in the

equilibrium state. The non-equilibrium state was achieved by refolding from chains, which were unfolded so quickly, that most of the *trans* configuration of the native state was maintained. Similar double jump experiments were performed in classical experiments with ribonuclease S [112]. The fast phase of collagen folding is biologically not very interesting, because natural collagens fold predominantly by the slow phase. The kinetics of folding of a collagen triple helix without restrictions by *cis-trans* isomerization is however an appealing biophysical questions. For the α-helix and the DNA double helix high folding rates were determined [113] but the maximum rate of collagen triple helix folding is still unknown.

As mentioned, collagen model peptides with fused oligomerization domains were used in several studies. As in true collagens nucleation is no longer rate limiting and again a fast and slow propagation reaction was observed. The rate constants derived for the slow phase match values expected for *cis-trans* isomerization (Table 4). An interesting study with such engineered collagen models was performed in order to solve the question, whether the kinetics of triple helix folding may be different for nucleation at the C- or N-terminus. The problem was approached by a comparison of model peptides $(GlyProPro)_{10}$, which were either linked to trimers at the N- or C-terminus [69]. Linkage was either achieved by fusion with a short segment containing the disulfide knot of collagen III or by attachment of the foldon domains. In both cases, a stabilization of the triple helix after cross-linking was observed. The kinetics of triple-helix formation was identical for all four model systems $(GlyProPro)_n$-foldon, foldon-$(GlyProPro)_n$, $(GlyProPro)_n$-Cys_2 and Cys_2-$(GlyProPro)_n$. Also the activation energies of folding were identical for all four peptides. Rate constants of folding were about 10^{-3} s^{-1} at 20 °C and the activation energy was 50 kJ/mol [69]. The data indicate that triple helix formation proceeds with closely similar rates in both directions. These findings are relevant in view of recent suggestions, that the folding may start at the N-terminus for a group of membrane bound collagens [47, 114].

11 Hysteresis

Unfolding and refolding of the triple helix is highly rate dependent in the transition region. Consequently, transition curves recorded by heating and by cooling form a hysteresis loop [115–117]. Hysteresis is very prominent for long natural collagens like collagen III and is also observed under isothermal conditions when unfolding is induced by increasing the concentration of guanidine hydrochloride. In the case of collagen III with an intact disulfide knot at the C-terminus, correct and complete refolding of native molecules is achieved at temperatures 10 or more degrees below the transition temperature (or 0.5 mol/l below the guanidinium hydrochloride midpoint concentration) after short incubation times of a few hours. In these cases, true hysteresis, in which equi-

librium values are not reached even after very long waiting times, is limited to the range near the transition temperature. As mentioned, unlinked single chains refold to misaligned structures at low temperatures. In these cases, an apparent hysteresis is observed, when monitoring circular dichroism or other parameters (Fig. 3). The nature of this apparent hysteresis is very different from the true hysteresis because refolding products differ from the native parent molecules.

It was noticed that hysteresis was less prominent for short collagen triple helices like the one-quarter fragment of collagen III [115]. More recently, it was found that hysteresis loops are clearly observable also for short model peptides as long as the heating and cooling rates are not too slow. The major difference between native long triple helices and short model peptides is apparently the time dependence. For long triple helices the loops persist even at very slow scan rates whereas for short triple helices equilibrium is achieved more quickly.

For the model systems with linked chains, a simple kinetic hysteresis mechanism is proposed (Boudko, Bächinger and Engel, in preparation). In the transition region the reciprocal apparent rate constant is

$$k_{app}^{-1} = \frac{1}{k_f + k_u} \tag{6}$$

in which k_f and k_u are the rate constants of folding and unfolding, respectively. Rate constant k_f is dependent on temperature according to the Arrhenius relation with the high activation energy E_a of *cis-trans* isomerization (see above). The temperature dependence of k_u is much lower and its activation energy can be calculated from and the enthalpy of the reaction. A satisfactory fit of the experimentally observed change of helicity F during heating or cooling is obtained when the rate law

$$\frac{dF}{dt} = k_f(1 - F) - k_u F \tag{7}$$

is integrated with the starting conditions (1) F=1 at t=0 for heating and (2) F=0 at t=0 for cooling. Case 1 yields the time course

$$F = \frac{K}{1 + K} + \frac{1}{1 + K} e^{-k_{app}t} \tag{8}$$

and case 2

$$F = \frac{K}{1 + K} (1 - e^{-k_{app}t}) \tag{9}$$

with $K=k_f/k_u$. The hysteresis loops can be constructed by calculation of F after different times at different temperatures. Obviously for infinite times t the equilibrium curve F=K/(1+K) is obtained.

The mechanism of the hysteresis for long triple helices of natural collagens is clearly more complicated. They do not fold in an all-or-none reaction as the short proteins as demonstrated by a cooperative length, which is about 1/10 of the length of the molecule [115]. A possible mechanism was proposed by Engel and Bächinger [81].

12 Mutations in Collagen Triple Helices Effect Proper Folding

Mutations in collagen genes cause a number of severe inherited diseases [118, 119]. The best-studied collagen related genetic disease is osteogenesis imperfecta (brittle bone disease). Point mutations at different positions of the triple helix lead to improperly folded and instable triple helices, disturbed fiber formation and severe pathological changes of the collagen matrix. Mutations near the C-terminus tend to be more severe than similar mutations near the N-terminus. In view of the steric need of small glycine residues in every third position (see above), mutations of glycines to residues with larger side chains are particularly disturbing and cause large decreases of transition temperatures. In some cases, even kinks were visualized in such mutated collagens [120]. A delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta was observed [121]. Hydroxylation of prolines only occurs in the unfolded state and consequently the extent of hydroxylation was much increased in the mutated collagens, apparently at the N-terminal side of the mutation. Recently model peptides were studied with sequence irregularities designed after important natural mutations [122]. With this approach, it is hoped to gain deeper explanations of how a single point mutation in the triple helix may cause a global pathological condition.

References

1. Rich A, Crick FH (1961) J Mol Biol 3:483
2. Ramachandran GN (1967) Structure of collagen at the molecular level. In: Ramachandran GN (ed) Treatise on collagen. Academic Press, New York, pp 103–183
3. Fraser RD, MacRae TP, Miller A, Suzuki E (1983) J Mol Biol 167:497
4. Golbik R, Eble JA, Ries A, Kuhn K (2000) J Mol Biol 297:501
5. Stetefeld J, Jenny M, Schulthess T, Landwehr R, Engel J, Kammerer RA (2000) Nat Struct Biol 7:772
6. Okuyama K, Arnott S, Takayanagi M, Kakudo M (1981) J Mol Biol 152:427
7. Bella J, Eaton M, Brodsky B, Berman HM (1994) Science 266:75
8. Nagarajan V, Kamitori S, Okuyama K (1999) J Biochem (Tokyo) 125:310
9. Kramer RZ, Bella J, Brodsky B, Berman HM (2001) J Mol Biol 311:131
10. Kramer RZ, Bella J, Mayville P, Brodsky B, Berman HM (1999) Nat Struct Biol 6:454
11. Bella J, Brodsky B, Berman HM (1995) Structure 3:893
12. Bella J, Brodsky B, Berman HM (1996) Connect Tissue Res 35:401
13. Kramer RZ, Vitagliano L, Bella J, Berisio R, Mazzarella L, Brodsky B, Zagari A, Berman HM (1998) J Mol Biol 280:623

14. Nagarajan V, Kamitori S, Okuyama K (1998) J Biochem (Tokyo) 124:1117
15. Hongo C, Nagarajan V, Noguchi K, Kamitori S, Okuyama K, Tanaka Y, Nishino N (2001) Polymer J 10:812
16. Berisio R, Vitagliano L, Mazzarella L, Zagari A (2002) Protein Sci 11:262
17. Vitagliano L, Berisio R, Mastrangelo A, Mazzarella L, Zagari A (2001) Protein Sci 10: 2627
18. Vitagliano L, Berisio R, Mazzarella L, Zagari A (2001) Biopolymers 58:459
19. Piez KA, Trus BL (1981) Biosci Rep 1:801
20. Bateman JF, Lamande SR, Ramshaw JAM (1996) In: Comper WD (ed) Extracellular matrix, vol 2. Harwood Academic Publishers, Amsterdam, p 27
21. Kühn K (1969) Essays Biochem 5:59
22. Chapman JA, Tzaphlidou M, Meek KM, Kadler KE (1990) Electron Microsc Rev 3:143
23. Daragan VA, Ilyina E, Mayo KH (1993) Biopolymers 33:5
24. Torchia DA, Lyerla JR Jr, Quattrone AJ (1975) Biochemistry 14:887
25. Fan P, Li MH, Brodsky B, Baum J (1993) Biochemistry 32:13299
26. Baum J, Brodsky B (2000) Case study 2. Folding of the collagen triple-helix and its naturally occurring mutants. In: Pain RH (ed) (series editors Hames BD, Glover DM) Frontiers in molecular biology: mechanisms of protein folding, 2nd edn. Oxford University Press
27. Li MH, Fan P, Brodsky B, Baum J (1993) Biochemistry 32:7377
28. Melacini G, Feng Y, Goodman M (1996) J Am Chem Soc 118:10359
29. Fiori S, Sacca B, Moroder L (2002) J Mol Biol 319:1235
30. Barth D, Kyrieleis O, Frank S, Renner C, Moroder L (2003) Chemistry 9:3703
31. Barth D, Musiol HJ, Schutt M, Fiori S, Milbradt AG, Renner C, Moroder L (2003) Chemistry 9:3692
32. Melacini G, Bonvin AM, Goodman M, Boelens R, Kaptein R (2000) J Mol Biol 300: 1041
33. Hofmann H, Voss T, Kuhn K, Engel J (1984) J Mol Biol 172:325
34. Vogel BE, Doelz R, Kadler KE, Hojima Y, Engel J, Prockop DJ (1988) J Biol Chem 263:19249
35. Kühn K (1995) Matrix Biol 14:439
36. Engel J (1997) Science 277:1785
37. Ricard-Blum S, Dublet B, van der Rest M (2000) Uncoventional collagens. Types VI, VII, VIII, IX, X, XIV, XVI and XIX. In: Sheterline P (ed) Uncoventional collagens. Types VI, VII, VIII, IX, X, XIV, XVI and XIX. Oxford University Press
38. Kishore U, Reid KB (1999) Immunopharmacology 42:15
39. Tschopp J (1982) Mol Immunol 19:651
40. Kishore U, Eggleton P, Reid KB (1997) Matrix Biol 15:583
41. Weis WI, Taylor ME, Drickamer K (1998) Immunol Rev 163:19
42. Wright JK, Tschopp J, Jaton JC, Engel J (1980) Biochem J 187:775
43. Krieger M, Herz J (1994) Annu Rev Biochem 63:601
44. McAlinden A, Smith TA, Sandell LJ, Ficheux D, Parry DA, Hulmes DJ (2003) J Biol Chem 278:42200
45. Lupas A (1996) Trends Biochem Sci 21:375
46. Kammerer RA (1997) Matrix Biol 15:555
47. Latvanlehto A, Snellman A, Tu H, Pihlajaniemi T (2003) J Biol Chem 278:37590
48. Sakakibara S, Kishida Y, Kikuchi Y, Sakai R, Kkiuchi K (1968) Bull Chem Soc Jpn 41:1273
49. Sakakibara S, Inouye K, Shudo K, Kishida Y, Kobayashi Y, Prockop DJ (1973) Biochim Biophys Acta 303:198
50. Suto K, Noda H (1974) Biopolymers 13:2477

51. Rothe M, Theyson R, Steffen KD, Schneider HJ, Zamani M, Kostrzewa M, Schindler W (1970) Angew Chem Int Ed Engl 9:535
52. Weber RW, Nitschmann H (1978) Helv Chim Acta 61:701
53. Shah NK, Ramshaw JA, Kirkpatrick A, Shah C, Brodsky B (1996) Biochemistry 35:10262
54. Kwak J, Jefferson EA, Bhumralkar M, Goodman M (1999) Bioorg Med Chem 7:153
55. Jenkins CL, Bretscher LE, Guzei IA, Raines RT (2003) J Am Chem Soc 125:6422
56. Heidemann E, Neiss HG, Khodadadeh K, Heymer G, Sheikh EM, Saygin O (1977) Polymer 18:420
57. Feng Y, Melacini G, Taulane JP, Goodman M (1996) Biopolymers 39:859
58. Fields CG, Lovdahl CM, Miles AJ, Hagen VL, Fields GB (1993) Biopolymers 33:16
59. Kwak J, De Capua A, Locardi E, Goodman M (2002) J Am Chem Soc 124:14085
60. Tanaka Y, Suzuki K, Tanaka T (1998) J Pept Res 51:413
61. Yu YC, Berndt P, Tirrell M, Fields GB (1996) J Am Chem Soc 118:12515
62. Yu YC, Tirrell M, Fields GB (1998) J Am Chem Soc 120:9979
63. Yu YC, Roontga V, Daragan VA, Mayo KH, Tirrell M, Fields GB (1999) Biochemistry 38:1659
64. Koide T, Yuguchi M, Kawakita M, Konno H (2002) J Am Chem Soc 124:9388
65. Mann K, Mechling DE, Bächinger HP, Eckerskorn C, Gaill F, Timpl R (1996) J Mol Biol 261:255
66. Mechling DE, Bächinger HP (2000) J Biol Chem 275:14532
67. Frank S, Kammerer RA, Mechling D, Schulthess T, Landwehr R, Bann J, Guo Y, Lustig A, Bächinger HP, Engel J (2001) J Mol Biol 308:1081
68. Boudko S, Frank S, Kammerer RA, Stetefeld J, Schulthess T, Landwehr R, Lustig A, Bächinger HP, Engel J (2002) J Mol Biol 317:459
69. Frank S, Boudko S, Mizuno K, Schulthess T, Engel J, Bächinger HP (2003) J Biol Chem 278:7747
70. Boudko SP, Engel J (2004) J Mol Biol 335:1289
71. Ottl J, Musiol HJ, Moroder L (1999) J Pept Sci 5:103
72. Sacca B, Renner C, Moroder L (2002) J Mol Biol 324:309
73. Engel J, Chen HT, Prockop DJ, Klump H (1977) Biopolymers 16:601
74. Privalov PL (1982) Adv Protein Chem 35:1
75. Gough CA, Bhatnagar RS (1999) J Biomol Struct Dyn 17:481
76. Doi M, Nishi Y, Uchiyama S, Nishiuchi Y, Nakazawa T, Ohkubo T, Kobayashi Y (2003) J Am Chem Soc 125:9922
77. Go N, Suezaki Y (1973) Letter: Analysis of the helix-coil transition in $(\text{Pro-Pro-Gly})_n$ by the all-or-none model. Biopolymers 12:1927
78. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B (2003) J Am Chem Soc 125:11500
79. Venugopal MG, Ramshaw JA, Braswell E, Zhu D, Brodsky B (1994) Biochemistry 33:7948
80. Bann JG, Bächinger HP (2000) J Biol Chem 275:24466
81. Engel J, Bächinger HP (2000) Matrix Biol 19:235
82. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B (2000) Biochemistry 39:14960
83. Chan VC, Ramshaw JA, Kirkpatrick A, Beck K, Brodsky B (1997) J Biol Chem 272:3144
84. Shah NK, Sharma M, Kirkpatrick A, Ramshaw JA, Brodsky B (1997) Biochemistry 36:5878
85. Engel J (1987) In: Advances in meat research, vol 4. van Nostrand Reinhold Company, New York, p 145
86. Dölz R, Heidemann E (1986) Biopolymers 25:1069
87. Bächinger HP, Davis JM (1991) Int J Biol Macromol 13:152
88. Stetefeld J, Jenny M, Schulthess T, Landwehr R, Schumacher B, Frank S, Rüegg MA, Engel J, Kammerer RA (2001) Nat Struct Biol 8:705

89. Pakkanen O, Hamalainen ER, Kivirikko KI, Myllyharju J (2003) J Biol Chem 278:32478
90. De Santis P, Giglio E, Liquori AM, Ripamonti A (1965) Nature 206:456
91. Dickerson RE, Geis I (1969) The structure and action of proteins. Benjamin, chap 2
92. Schulz GE, Schirmer RH (1979) Principles of protein structure. In: Cantor CR (ed) Springer advanced texts in chemistry. Springer, Berlin Heidelberg New York Tokyo
93. Burjanadze TV (1979) Biopolymers 18:931
94. Inouye K, Kobayashi Y, Kyogoku Y, Kishida Y, Sakakibara S, Prockop DJ (1982) Arch Biochem Biophys 219:198
95. Inouye K, Sakakibara S, Prockop DJ (1976) Biochim Biophys Acta 420:133
96. Mizuno K, Hayashi T, Peyton DH, Bächinger HP (2004) J Biol Chem 279:282
97. Engel J, Prockop DJ (1998) Matrix Biol 17:679
98. Holmgren SK, Taylor KM, Bretscher LE, Raines RT (1998) Nature 392:666
99. Bretscher LE, Jenkins CL, Taylor KM, DeRider ML, Raines RT (2001) J Am Chem Soc 123:777
100. Raines RT, Bretscher LE, Holmgren SK, Taylor KM (1999). The stereoelectronic basis of collagen stability. In: Fields GB, Tam JP, Barany G (eds) Peptides for the new millenium. Proceedings of the 16th American Peptide Symposium. Kluwer Academic, Norwell, MA
101. Renner C, Alefelder S, Bae JH, Budisa N, Huber R, Moroder L (2001) Angew Chem Int Ed Engl 40:923
102. Scherer G, Kramer ML, Schutkowski MUR, Fischer G (1998) J Am Chem Soc 120:5568
103. Gratwohl C, Wüthrich K (1976) Biopolymers 15:2025
104. Reimer U, Scherer G, Drewello M, Kruber S, Schutkowski M, Fischer G (1998) J Mol Biol 279:449
105. Sarkar SK, Young PE, Sullivan CE, Torchia DA (1984) Proc Natl Acad Sci USA 81:4800
106. Xu Y, Bhate M, Brodsky B (2002) Biochemistry 41:8143
107. Bächinger HP, Bruckner P, Timpl R, Engel J (1978) Eur J Biochem 90:605
108. Dölz R, Engel J, Kühn K (1988) Eur J Biochem 178:357
109. Bruckner P, Bächinger HP, Timpl R, Engel J (1978) Eur J Biochem 90:595
110. Bächinger HP, Bruckner P, Timpl R, Prockop DJ, Engel J (1980) Eur J Biochem 106:619
111. Bruckner P, Eikenberry EF (1984) Eur J Biochem 140:391
112. Brandts JF, Halvorson HR, Brennan M (1975) Biochemistry 14:4953
113. Schwarz G, Engel J (1972) Angew Chem Int Ed Engl 11:568
114. Franzke CW, Tasanen K, Schumann H, Bruckner-Tuderman L (2003) Matrix Biol 22:299
115. Davis JM, Bachinger HP (1993) J Biol Chem 268:25965
116. Bächinger HP, Engel J (2001) Matrix Biol 20:267
117. Leikina E, Mertts MV, Kuznetsova N, Leikin S (2002) Proc Natl Acad Sci USA 99:1314
118. Kuivaniemi H, Tromp G, Prockop DJ (1991) Faseb J 5:2052
119. Uitto J, Lichtenstein JR (1976) J Invest Dermatol 66:59
120. Engel J, Prockop DJ (1991) Annu Rev Biophys Biophys Chem 20:137
121. Raghunath M, Bruckner P, Steinmann B (1994) J Mol Biol 236:940
122. Baum J, Brodsky B (1999) Curr Opin Struct Biol 9:122

Top Curr Chem (2005) 247: 35–84
DOI 10.1007/b103819
© Springer-Verlag Berlin Heidelberg 2005

The Collagen Superfamily

Sylvie Ricard-Blum[1(✉),4] · Florence Ruggiero[2] · Michel van der Rest[3]

[1] Institut de Biologie Structurale, CEA-CNRS UMR 5075, Université Joseph Fourier
41 rue Jules Horowitz, 38027 Grenoble Cedex 01, France
ricard@ibs.fr
[2] Institut de Biologie et Chimie des protéines, CNRS UMR 5086, IFR128 Biosciences Gerland,
7 passage du Vercors, 69367 Lyon Cedex 07, France
f.ruggiero@ibcp.fr
[3] Ecole Nationale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France
Michel.Van.Der.Rest@ens-lyon.fr
[4] Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS-UCBL
7, passage du Vercors, 69367 Lyon Cedex 07, France (Present address)

1 Introduction . . . 37

2 The Collagen Family: an Update . . . 38
2.1 Fibrillar Collagens . . . 49
2.1.1 Features of Fibrillar Collagens . . . 49
2.1.2 The Novel Fibrillar Collagens: Collagens XXIV and XXVII . . . 51
2.1.3 New Insights on the Classical Collagen V Structure and Functions . . . 52
2.2 FACITs (Fibrillar-Associated Collagens with Interrupted Triple-Helix) . . . 53
2.2.1 Recent Advances on Collagens IX, XII and XIV . . . 53
2.2.2 Structure of the Novel FACITS: Collagens XVI, XIX, XX, XXI and XXII . . . 54
2.2.3 Distribution and Function of the Novel FACITs . . . 55
2.3 Membrane Collagens and Collagen-Like Proteins . . . 57
2.3.1 Common Features . . . 57
2.3.2 Membrane Collagens . . . 58
2.3.3 The Membrane Collagen-Like Proteins . . . 59
2.4 Multiplexins . . . 60
2.4.1 Common Features . . . 60
2.4.2 Collagen XV . . . 61
2.4.3 Collagen XVIII . . . 61
2.5 Collagen XXVI and the Emu Family . . . 62
2.5.1 Emilin1 and Emilin2 . . . 62
2.5.2 Emilin and Multimerin-Domain Containing Proteins (Emu1 and Emu2) . . . 62
2.5.3 Collagen XXVI is Identical to the Emu2 Protein . . . 63
2.6 Collagen-Like Molecules . . . 63

3 Non Collagenous Domains of Collagens . . . 64
3.1 Cystein-Rich Repeat Domain . . . 65
3.2 Thrombospondin-1 Domain . . . 65
3.3 Kunitz Domain of the α3(VI) Chain . . . 66
3.4 NC1 Domains of Collagen IV . . . 66
3.5 NC1 Domains of Network-Forming Collagens (VIII and X) . . . 68
3.6 Endostatins XVIII and XV . . . 69

4 Collagen-Related Diseases and Prospects for Gene Therapy 71
4.1 Osteogenesis Imperfecta 72
4.2 Bethlem Myopathy 72
4.3 Alport Syndrome 72
4.4 Epidermolysis Bullosa 73
4.5 Corneal Endothelial Dystrophies 74
4.6 Schmid Metaphyseal Chondrodysplasia 74

5 Collagens and Neurodegenerative Diseases 75

6 Conclusion 76

References 78

Abstract The collagen family is highly complex and shows a remarkable diversity in molecular and supramolecular organization, tissue distribution and function. Collagen types are classified in several sub-families according to sequence homologies and to similarities in their structural organization and supramolecular assembly. This overview covers advances made in this field during the last five years. Due to space limitation, the reader is referred to previous reviews for the topics not discussed hereafter. We focus on recently described fibrillar collagens (collagens XXIV and XXVII) and FACITs (collagens XVI, XIX, XX, XXI and XXII), multiplexins (collagens XV and XVIII), membrane-collagens (collagens XIII, XVII, XXIII and XXV) and collagen XXVI, on other proteins containing triple-helical domains including the members of the new Emu family and on the structure and functions of several non collagenous domains found in collagens. We also discuss data on collagen-related diseases with particular emphasis on gene therapy and on the involvement of collagens in neurodegenerative diseases, which emerge as a major threat for public health in aging populations.

Keywords Collagen · Collagen-related diseases · Extracellular domains

List of Abbreviations

BMP	Bone Morphogenetic Protein
COL	Collagenous domain
CQT5	C1q tumor necrosis factor-related protein 5
CRR	Cystein-Rich Repeat domain
DDR	Discoïdin Domain Receptors
EMI	N-Terminal cysteine-rich domain found in the Emu family
Emu	Emilin and multimerin
ERK	Extracellular-signal-Regulated Kinase
FACIT	Fibril-Associated Collagen with Interrupted Triple Helix
FN3	Fibronectin type III repeat
FAK	Focal Adhesion Kinase
MARCO	MAcrophage Receptor with COllagenous structure
NC	Non-Collagenous domain
PKC	Protein Kinase C
RDEB	Recessive Dystrophic Epidermolysis bullosa
TSPN	Thrombospondin N-terminal like domain
SR-AI, SR-AII	Scavenger receptors type A

TGF-β	Transforming Growth Factor β
TNF	Tumor Necrosis Factor
vWA	von Willlebrand factor A-like domain

1 Introduction

Collagens are trimeric molecules made of three polypeptide α chains, which contain the sequence repeat (G-X-Y)n, X being frequently proline and Y hydroxyproline. These repeats allow the formation of a triple helix, which is the characteristic feature of the collagen superfamily. The side chains of each X and Y residue are at the surface of the triple helix, giving the collagen molecule a significant capacity for lateral interactions with other molecules of the extracellular matrix and resulting in the formation of various supramolecular assemblies. Collagens are modular proteins containing not only triple-helical (COL) but also non triple helical (NC) domains found in a number of other extracellular matrix proteins [1, 2]. Since many collagen chains were initially characterized by partial sequencing missing the sequences encoding the N-terminus of the protein, investigators in the field numbered these domains starting from the C-terminus of the protein. The reverse order has however been used by others. This is the case for collagens VII [3], XIII [4], XXII [5], XXIII [6], XXV [7] and XXVI [8].

Collagen nomenclature is somewhat confusing. Although there are no official criteria, it is generally considered that, for a protein to be called a collagen, it should contain at least one triple helical domain, should be able to form supramolecular aggregates and should be deposited within the extracellular matrix. Most of the members of the collagen family fulfill these criteria. However, collagens XIII, XVII, XXIII and XXV contain a transmembrane domain and do not appear to form supramolecular structures of their own, but they contain collagenous domains located within the extracellular matrix. Multiplexins (*multiple* triple heli*x* domains and *in*terruptions) including collagens XV and XVIII contain a triple helical domain with multiple interruptions and are found in basement membrane zones, but their supramolecular structures are still unknown. On the other hand, some molecules such as emilins, which contain a collagenous domain and are associated with supramolecular assemblies (elastic fibers) in the extracellular matrix are not classified as collagens at this time.

With the completion of the sequencing of the human genome, it is likely that most human collagen genes are now discovered. The first collagens to be studied were extracted from tissues and characterized at the protein level by various biochemical methods (Fig. 1). Then, most of the newly discovered collagens were first identified as cDNA clones. One of the most recently described collagen, collagen XXVI, was identified by yeast two-hybrid screening of a whole mouse embryo cDNA library using the collagen-specific molecular chaperone HSP47 as a bait [8].

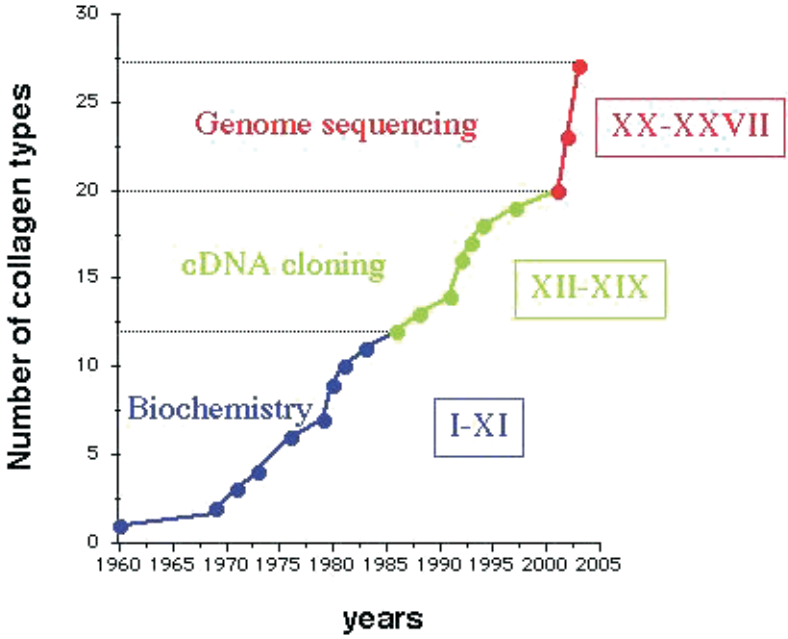

Fig. 1 Discovery of the collagen super family members: a 50-year story

2 The Collagen Family: an Update

For a comprehensive coverage of the structure and functions of collagens I–XIV, which are not discussed in this issue, the reader is referred to detailed reviews on fibrillar collagens [9] and unconventional collagens including collagen VII, network-forming collagens (VI, VIII and X) and the FACITs (IX, XII, XIV, XVI and XIX) [10] and to recent general reviews [11, 12]. Due to space limitation, sequence data, splicing variants and domain organization of each collagen chain will not be detailed below. Swiss Prot or TrEMBL access numbers for all the members of the collagen superfamily (Tables 1 and 2) are provided for easy access of the reader to these data. Intermolecular interactions of collagens with other constituents of extracellular matrix and with cells fall outside the scope of this overview and the reader is referred to the Protein Profile Series dedicated to fibrillar [9] and unconventional collagens [10] for full report on these topics. This section will focus on the most recently described collagens and on proteins containing triple-helical domains, the number of which has also increased in the past years.

Table 1 Swiss Prot or TrEMBL accession number and entry name in the Swiss Prot database (http://us.expasy.org/sprot/) are provided for collagen alpha chains. Entries are provided for human proteins, unless otherwise stated

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Collagen α1(I) chain	CA11_HUMAN	P02452
Collagen α2(I) chain	CA21_HUMAN	P08123
Collagen α1(II) chain	CA12_HUMAN	P02458
Collagen α1(III) chain	CA13_HUMAN	P02461
Collagen α1(IV) chain	CA14_HUMAN	P02462
Collagen α2(IV) chain	CA24_HUMAN	P08572
Collagen α3(IV) chain	CA34_HUMAN	Q01955
Collagen α4(IV) chain	CA44_HUMAN	P53420
Collagen α5(IV) chain	CA54_HUMAN	P29400

Table 1 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Collagen α6(IV) chain	CA64_HUMAN	Q14031
Collagen α1(V) chain	CA15_HUMAN	P20908
Collagen α2(V) chain	CA25_HUMAN	P05997
Collagen α3(V) chain	CA35_HUMAN	P25940
α4 type V collagen (rat)	Q9JI04	Q9JI04 (TrEMBL)
Collagen α1(VI) chain	CA16_HUMAN	P12109
Collagen α2(VI) chain	CA26_HUMAN	P12110
Collagen α3(VI) chain	CA36_HUMAN	P12111
Collagen α1(VII) chain	CA17_HUMAN	Q02388

Table 1 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Collagen α1(VIII) chain	CA18_HUMAN	P27658
Collagen α2(VIII) chain [Fragment]	CA28_HUMAN	P25067
Collagen α1(IX) chain	CA19_HUMAN	P20849
Collagen α2(IX) chain	CA29_HUMAN	Q14055
Collagen α3(IX) chain	CA39_HUMAN	Q14050
Collagen α1(X) chain	CA1A_HUMAN	Q03692
Collagen α1(XI) chain	CA1B_HUMAN	P12107
Collagen α2(XI) chain	CA2B_HUMAN	P13942
Collagen α1(XII) chain	CA1C_HUMAN	Q99715

Table 1 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
α1 type XIII collagen	Q14035	Q14035 (TrEMBL)
Collagen type XIV (mouse)	Q80X19	Q80X19 (TrEMBL)
Collagen alpha 1(XV) chain	CA1E_HUMAN	P39059
Collagen alpha 1(XVI) chain	CA1F_HUMAN	Q07092
– Collagen, type XVII, alpha 1 (BP180) or BA16H23.2	Q9NQK9	Q9NQK9 (TrEMBL)
– 180 kDa bullous pemphigoid antigen 2/type XVII collagen	Q9UMD9	Q9UMD9 (TrEMBL)

Table 1 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Collagen alpha 1(XVIII) chain	CA1H_HUMAN	P39060
Collagen alpha 1(XIX) chain	CA1I_HUMAN	Q14993
Collagen type XX alpha 1 (Chicken)	Q90ZA0	Q90ZA0 (TrEMBL)
Alpha 1 type XXI collagen precursor	Q96P44	Q96P44 (TrEMBL)
Alpha 1 type XXII collagen	Q8NFW1	Q8NFW1 (TrEMBL)
Alpha 1 type XXIII collagen	Q86Y22	Q86Y22 (TrEMBL)
Alpha 1 type XXIV collagen	Q7Z5L5	Q7Z5L5 (TrEMBL)

Table 1 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Collagen-like Alzheimer amyloid plaque component precursor type I (Mouse)	Q99MQ5	Q99MQ5 (TrEMBL)
Collagen alpha 1(XXVI) chain (Emu2 protein, Emilin and multimerin-domain containing protein 2)	EMU2_HUMAN	Q96A83
Collagen XXVII proα1 chain	Q8IZC6	Q8IZC6 (TrEMBL)

Table 2 Swiss Prot or TrEMBL accession number and entry name in the Swiss Prot database (http://us.expasy.org/sprot/) are provided for collagen-like molecules. Entries are provided for human proteins, unless otherwise stated

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Chipmunk hibernation-related proteins		
– Hibernation-associated plasma protein HP-20	HP20_TAMSI	Q06575
– Hibernation-associated plasma protein HP-25	HP25_TAMSI	Q06576
– Hibernation-associated plasma protein HP-27	HP27_TAMSI	Q06577
Saccular collagen (Inner ear-specific collagen) (Bluegill)	COLE_LEPMA	P98085
Acetylcholinesterase collagenic tail peptide (AChE Q subunit)	COLQ_HUMAN	Q9Y215
Complement C1q subcomponent		
– A chain	C1QA_HUMAN	P02745
– B chain	C1QC_HUMAN	P02746
– C chain	C1QB_HUMAN	P02747
Complement-c1q tumor necrosis factor-related protein 5	CQT5-Human	Q9BXJ0

Table 2 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
C1q-related factor	C1RF_HUMAN	O75973
Adiponectin (30 kDa adipocyte complement-related protein)	APM1_HUMAN	Q15848
Emilin-1 (Elastin microfibril interface-located protein 1)	EMI1_HUMAN	Q9Y6C2
Emilin-2 (Elastin microfibril interface-located protein 2)	EMI2_HUMAN	Q9BXX0
Emu1 protein (Emilin and multimerin-domain containing protein 1)	EMU1_HUMAN	Q96A84
Ectodysplasin A	EDA_HUMAN	Q92838
Macrophage scavenger receptor types I and II	MSRE_HUMAN	P21757–1 P21757–2

Table 2 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Scavenger receptor with C-type lectin type I (SRCL)	Q9BYH7	Q9BYH7 (TrEMBL)
Scavenger receptor with C-type lectin type II (SRCL)	Q9BY85	Q9BY85 (TrEMBL)
Macrophage receptor MARCO (Macrophage receptor with collagenous structure)	MRCO_HUMAN	Q9UEW3
Collectin placenta 1 (CL-P1)	Q8WZA4	Q8WZA4 (TrEMBL)
Collectin 34	Q9Y6Z7	Q9Y6Z7 (TrEMBL)
Pulmonary surfactant-associated protein A (SP-A)	PSPA_HUMAN	P07714
Pulmonary surfactant-associated protein D (SP-D)	PSPD_HUMAN	P35247

Table 2 (continued)

Protein name	Entry name in Swiss prot or TrEMBL	Swiss Prot Primary Accession Number
Mannose-binding protein C	MABC_HUMAN	P11226
Ficolin 1 (Ficolin A, M-Ficolin)	FCN1_HUMAN	O00602
Ficolin 2 (Ficolin B, L-Ficolin, serum lectin p35)	FCN2_HUMAN	Q15485
Ficolin 3 (Collagen/fibrinogen domain-containing lectin 3 p35)	FCN3_HUMAN	O75636

2.1 Fibrillar Collagens

2.1.1 Features of Fibrillar Collagens

The fibril-forming or fibrillar collagens represent the most abundant product synthesized by connective tissue cells. As a consequence, they have also been the first members of the collagen superfamily to be discovered. Since 1979, the year of collagen XI discovery, the fibrillar collagens comprised five members: the quantitatively major fibrillar collagens types I, II and III, and the minor collagen types V and XI [13]. As no novel fibrillar collagens has been described until very recently, the term of "classical" collagens is also commonly used to classify this subset of the collagen family. Fibrillar collagens all share a common chain structure composed of a large collagenous domain (COL1) bordered by the N- and C-terminal extensions called the N- and C-propeptides, respectively. The C-propeptide is referred to as the NC1 domain, the N-propeptide is divided into sub-domains: a short sequence (NC2) that links the major triple helix to the minor one (COL2), and a globular N-terminal end (NC3) that can show structural variations and splicing alternatives (Fig. 2A). Notably, alternative splicing within the NC3 domain of proα1(II) mRNA generates a long form of collagen II, the collagen IIA and a short form, IIB with distinct tissue localization and function as discussed lower than (see below). Fibrillar collagens are synthesized as precursors containing both propeptides, referred to as procollagens, that are secreted in the extracellular space. Fibrils represent the final product of classical fibrillar collagen biosynthesis. Thus far, only fibrillar procollagens require proteolytic removal of the N and C-terminal extensions to achieve a mature and functional molecule. This step was considered as an absolute prerequisite for correct fibril formation (see [14, 15] for reviews). This dogma was revisited with the observation that the mature molecule of minor fibrillar collagens V and XI conserved a part of their N-terminal extension. The persistence of the N-propeptide is in part responsible for the control of heterotypic growth by sterically limiting lateral molecule addition (see [13] for review). The N-propeptide of the classical fibrillar collagen represents the domain that shows the most important sequence and structure variations between the different members. Nevertheless, almost all fibrillar collagen chains can be divided into two separate groups according to the structure of their N-terminal globular domain referred to as the NC3 domain (Fig. 2A). The NC3 domains of α1(I), α1(IIA), α1(III) and α2(V) chains all harbor a cystein-rich repeat domain (CRR) and will be designed hereafter as the CRR-containing chains, while a thrombospondin N-terminal-like (TSPN) domain is found in α1(V), α3/α4(V), α1(XI), α2(XI), α1(XXIV) and α1(XXVII). New interesting data have been reported on the structure and functions of these two domains that will be presented below.

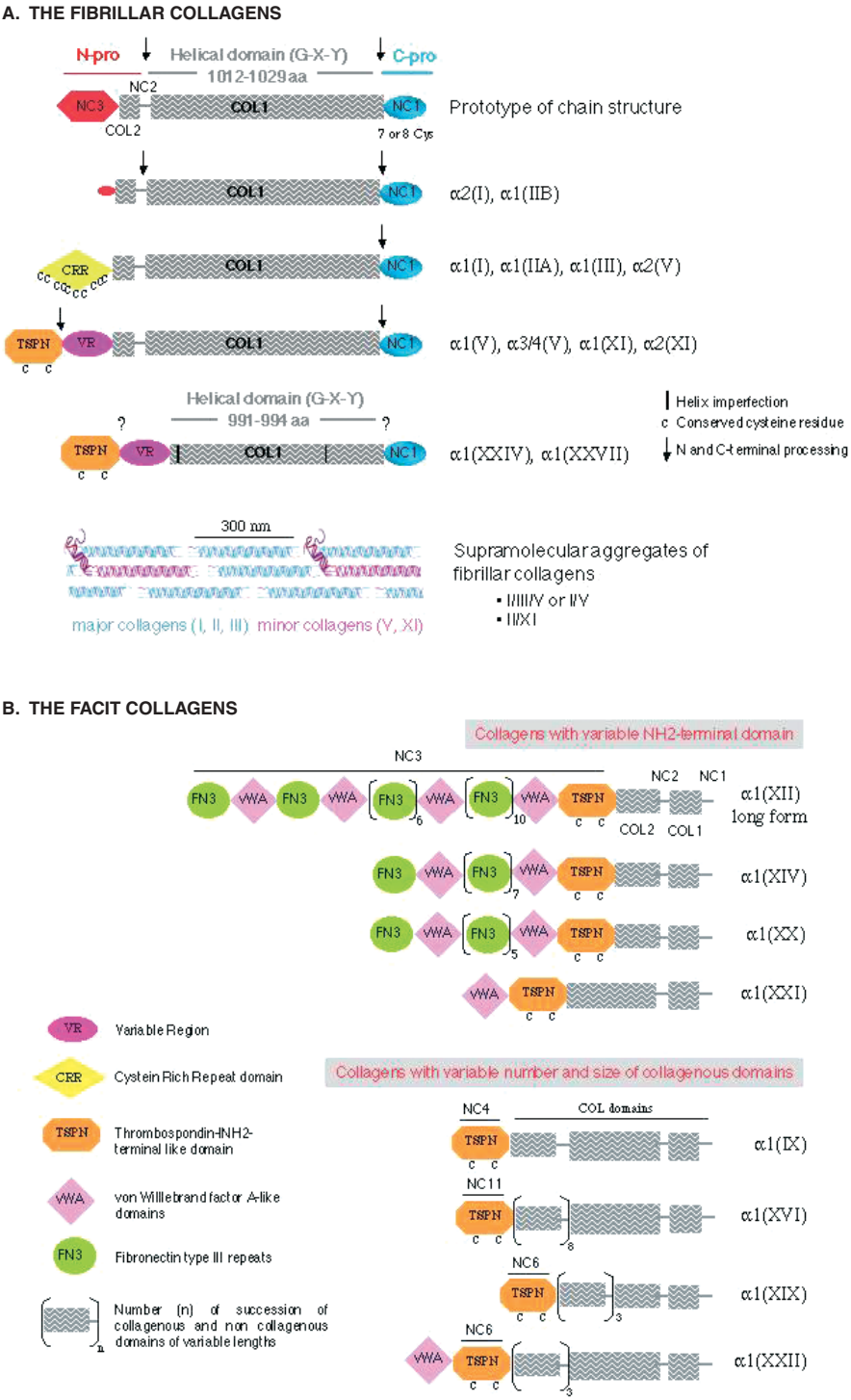

Fig. 2A, B Fibrillar collagens and FACITS

2.1.2
The Novel Fibrillar Collagens: Collagens XXIV and XXVII

Unexpectedly, last year, two novel members of the fibrillar collagen group, numbered collagens XXIV [16] and XXVII [17, 18] have been identified. They both share the structural features of fibrillar collagens, namely a long collagenous domain flanked at both extremities by globular N- and C-propeptides. Their large amino-terminal domains (up to 550 residues) are closely related to those of collagen V and XI that comprise two sub-domains, the thrombospondin N-terminal-like motif, referred to as TSPN, and a variable region (Fig. 2A). The C-propeptide is well conserved and presents 8 cysteine residues suggesting a possible homotrimeric association of these collagens. However, these two novel collagens present several features unusual for mammalian collagens, but shared with invertebrate fibrillar collagens [19]. The major triple helix differs substantially from classical fibrillar collagens. It is shorter than in the equivalent domain of the classical collagens and they contain imperfections in the $(G\text{-}X\text{-}Y)_n$ repeats. The significance of triple helix imperfections is not clear and biochemical investigation is required to elucidate their physiological relevance. Minor interruptions of the G-X-Y repeats in the major triple helical domain of the classical fibrillar collagens are known to cause disease [15]. However, triple helix imperfections resulting in kinks along the rigid rod-like molecular structure might also have biological implications. These particular flexible regions might hold potential binding sites accessible for molecular and cellular interactions or for protease activity. The total length of the major triple helix of the novel collagens varies from 991 to 997 amino acids depending on the collagen chains and species, but is always shorter than the equivalent fibrillar collagen domains (1014–1020 residues). However, the exact number of residues of the collagenous domain is debatable and may vary depending on the inclusion or not of a short triple helical sequence found in these collagens in close proximity to the main triple helical domain. Based on the structural features of invertebrate fibrillar collagens [20], the first N-terminal triple helix interruption might be considered as a break between the minor and the major triple helix. The resulting minor triple helix would then consist of 19 triplets and the major triple helix would then be 931 residues in length. Human collagen XXVII also contains unexpected residues such as tryptophan and cysteine within the triple helix domain. There significance is not clear since they are not all conserved among species and in the closely related collagen XXIV.

Classical fibrillar collagens share the unique property to aggregate into highly-ordered fibrils. Most of the collagen fibrils are made up of different fibrillar collagen types and are referred to as heterotypic fibrils (Fig. 2A). The current view of their tissue distribution is that heterotypic fibrils composed of collagen I, III and/or V are widely distributed in connective tissues whereas cartilaginous tissues and vitreous contain collagen II/XI mixed fibrils. Whether the novel collagens can incorporate heterotypic banded fibrils is not known. Based on the known fibrillogenesis mechanisms, the unusual structural fea-

tures of these collagens such as the distortion and reduced length of the triple helical domain argue against this possibility.

Unlike the collagen V heterotrimer, collagen V homotrimer that presents a kink within the triple helix generated by the presence of an imino-acid poor region does not incorporate into collagen I fibrils [21, 22]. The two "imperfect" fibrillar collagens might interfere with fibril formation and thus act as a negative regulating factor of heterotypic fibril growth. The distinct expression patterns described for *col24a1* and *col27a1* genes indicate that a close association of collagen XXVII with collagens II and XI, and alternatively of collagen XXIV with collagens I and V is likely to occur. Collagen XXIV expression was found strictly confined to bone and cornea both containing heterotypic fibrils formed with collagens I and V solely, and is thus not found in collagen III containing tissues such as skin, tendon and vessels. Collagen XXVII exhibited a strong expression in cartilage and, for this reason, represents a new member of the cartilaginous subset of the fibrillar collagen family.

2.1.3 New Insights on the Classical Collagen V Structure and Functions

Although the biosynthesis of classical fibrillar collagen has been extensively studied since their early discovery, some questions regarding their structure, processing and fibril formation remain astonishingly to be answered. Some of them have been elucidated during the last five years and are reviewed by others in this issue, notably concerning enzymes ensuring the processing to mature molecule and the molecular mechanisms that drive self-assembly of monomers. Another unexpected recent result came from collagen V chain studies. Apart from the most abundant form found in tissues, the heterotrimer [α1(V)]2α2(V), several other different chain associations exist for collagen V including hybrid molecules formed with both collagen V and XI chains, and strongly suggest a tissue-specific functions for collagen V. The heterotrimeric form, α1(V)α2(V)α3(V), only found in placenta [23], remains poorly studied until the recent determination of the primary structure of the human α3(V) chain, which will henceforth facilitate future research [24]. The homotrimeric form [α1(V)]3 was first observed in cell cultures and is likely to be present in embryonic tissues, though in too small amounts to be purified. Recently, however, using recombinant technology, the molecular characterization of the collagen V homotrimer and its role in fibril formation have been achieved [21, 22].

The diversity of the collagen V molecular forms has been recently augmented with the identification of a novel member of collagen V gene family referred to as α4(V) [25]. Whether this chain identified in rat does represent a novel chain or corresponds to an orthologue of the mouse and human α3(V) chain is not clearly established yet. The rat proα4(V) chain presents a very high percentage of identity with the human proα3(V) chain. In fact, the α1(V), α3(V), α4(V), α1(XI) and α2(XI) are closely related and were proposed to

constitute a separate clade among the fibrillar collagen genes [19]. However, from the current data, α3(V) chain and α4(V) chains differ by several features. Whereas the α3(V) chain is present in placenta as the heterotrimer α1(V)-α2(V)α3(V), the α4(V)-containing molecule shows a different heterotrimeric chain association [α1(V)]2α4(V)]. Moreover, proα4(V) chain presents a unique expression pattern. It was shown to be solely synthesized by Schwann cells and its expression is restricted to embryonic and early postnatal peripheral nerves. The α4(V)-collagen exhibits particular high affinity ($K_D \sim 0.2$ nmol/l) for the heparan sulfate transmembrane proteoglycan syndecan 3 [26]. Most importantly, this interaction with the cells is mediated by an heparin binding site located in the variable region of the N-propeptide, a finding likely important for its biological function. This flexible domain that persists on the mature molecule is known to be accessible at the heterotypic fibril surface and thus, unlike the triple helix probably buried in the fibrils, could interact with receptors at the cell surface. A number of collagen V receptors have been already identified [27–30], but they all bind to the collagen V major triple helix. Other isoforms bind heparin and heparan sulfate, but the binding site was localized within the triple helical domain of the α1(V) chain [31, 32]. In addition, the α4(V)-collagen was shown to display domain-specific effects that regulate peripheral nerves development. The α4(V) N-propeptide recombinantly expressed as a monomer was shown to stimulate Schwann cell migration through its heparin binding site whereas the major triple helix blocked neurite outgrowth [26]. The interaction through the N-propeptide domain was shown to induce actin cytoskeleton assembly, tyrosine phosphorylation and activation of the Erk1/Erk2 protein kinases [33].

2.2
FACITs (Fibrillar-Associated Collagens with Interrupted Triple-Helix)

2.2.1
Recent Advances on Collagens IX, XII and XIV

Beside a common general chain structure, the name of FACIT comes from the fact that the first discovered members of this subfamily, collagens IX, XII and XIV were shown to interact with collagen fibrils. The lateral interaction of collagen IX at the fibril surface through covalent crosslinks was proposed to act as a molecular bridge between cartilage fibrils and other matrix components. Recent advances on structural and functional aspects of this cartilaginous FACIT have been reviewed [34]. Although collagens XII and XIV localized at the fibril surface of collagen I- or II-containing fibrils, the detailed molecular mechanisms of the interactions remain to be clarified [35, 36]. No covalent cross-linking involving these collagen types could be demonstrated and the current point of view is that collagens XII and XIV might interact with fibrils through the matrix small proteoglycans decorin and fibromodulin [37–39]. Functions of collagens XII and XIV have not been fully elucidated but a line of

evidence puts forward a role in stabilizing and/or in organizing fibrillar network in extracellular matrix. Collagen XIV was reported to promote gel contraction [40] and might act as a regulating factor of fibril growth in developing tendon [41]. Collagen XII expression is prominently induced in response to wounding caused by excess mechanical stress. Its production was three- to fivefold higher in fibroblasts cultured in tensed than in released collagen type-I gels. The intracellular mechanisms controlling mechano-dependent production of the collagen XII required activation of the FAK-Erk1/2 pathway and of a distinct classical or novel PKC. Fibronectin was also up-regulated in cells under tensile stress but its expression induction is regulated by a different PKC-dependent pathway [42].

2.2.2 Structure of the Novel FACITS: Collagens XVI, XIX, XX, XXI and XXII

Recently, the FACIT sub-class of the collagen family has expanded to 8 members and henceforth includes the novel collagens XVI, XIX, XX, XXI and the most recently reported collagen XXII [5]. Although their size and their primary structure vary greatly, they all share the general structural features of the previously identified FACITs. The key features of this subfamily are signed by two highly conserved cysteine residues separated by four amino acids at the NC1-COL1 junctions and the presence of two G-X-Y triplet imperfections within the COL1 domain. They also display tremendous divergence in size (from 60 kDa to 220 kDa) particularly for their N-terminal NC domain. In addition, alternative splicing of collagens XII, XIV and XIX mRNAs and possibly of collagen XX occurs in some species, making their functional studies a difficult task to tackle (see [10] for review). Two (collagens XII, XIV, XX and XXI) or more (collagens IX, XVI, XIX and XXII) short triple helical domains are connected and bordered by non collagenous domains (Fig. 2B). The N-terminal non-collagenous domain diverges significantly among the different FACIT members. As indicated in the introduction and in Fig. 2B, collagen domains are commonly numbered from the C-terminus to the N-terminus. However this convention is not always respected and the reverse order can also be found in the literature, complicating the understanding of non-specialist readers. Since the number of collagenous domains in FACIT collagen varies, the N-terminal domains correspond to NC3 in collagens XII, XIV, XX, XXI, to NC4 in collagens IX, NC6 in collagens XIX and to NC11 in collagen XVI that contains the highest number of collagenous domains. FACITs usually contain a TSPN domain next to the collagenous domain. It represents the sole module constituting the N-terminal non collagenous domain of collagens IX, XVI and XIX.

Additional domains with structural homology to modules found in other molecules constitute the N-terminal domain of other FACITs. In collagens XII, XIV and XX, von Willlebrand factor A-like domains (vWA) alternate with fibronectin type III repeats (FN3), while in collagens XXI and XXII a unique vWA domain is placed next to the TSPN domain (Fig. 2B). When particularly

large as in collagens XII and XIV, the N-terminal domain folds into a trident-like structure as observed by rotary shadowing whereas N-terminal domains of collagens IX or XIX formed a large compact globular domain (reviewed in [10, 43]). Due to their numerous non-collagenous domains, FACITs are considered to be flexible molecules. This is supported by rotary shadowing images of the molecules that show several kinks interrupting the rod-like collagenous domains [5, 43].

2.2.3 Distribution and Function of the Novel FACITs

Although some FACITs, notably collagens XVI [44, 45] and XIX [46], were discovered more than ten years ago, very little is known about the role of the novel FACITs in connective tissues. However, the establishment of their expression patterns was in some cases instructive and could further help in elucidating their functions. In the absence of strong evidence for specific functions, FACITs are usually considered to be implicated in the stabilization and integrity of the extracellular matrices.

2.2.3.1 Collagen XVI

Collagen XVI is widely distributed in embryonic and adult tissues [47] but its supramolecular organization was reported to be tissue-specific [48]. In cartilage, collagen XVI resides in thin fibrils that contain also collagens II and XI, an expected localization for a member of the fibril-associated collagen family. In skin however, collagen XVI occurs at the dermo-epidermal interface in close vicinity to collagen VII, that forms the anchoring fibrils, and in the papillary dermis where it co-localizes with fibrillin-1. It is almost absent from the deeper layer, the reticular dermis [48–50]. Dermal fibroblasts, and to a lesser extent keratinocytes, produce substantial amounts of collagen XVI [50]. Collagen XVI is synthesized as a homotrimer and does not undergo further processing [51]. Its expression is functionally regulated. It is enhanced during cell growth arrest [52] and in fibrotic diseases [49]. Unlike the situation in normal tissues, fibroblasts actively migrate and proliferate in fibrotic conditions. Thus, collagen XVI was proposed to play a role in stabilizing fibroblasts in dermal matrices.

2.2.3.2 Collagen XIX

Based on its structural features and chain organization, collagen XIX belongs undoubtedly to the FACIT family. It appears more closely related to collagen XVI. However its preferential localization in basement membrane zones remains unique for a FACIT member. An homologue of this evolutionary conserved

collagen was identified as an adult-specific collagen of *C. elegans* cuticle [53]. Originally isolated from a human rhabdomyosarcoma cell line [46, 54], collagen XIX localizes in endothelial, neuronal, mesenchymal, and most epithelial basement membrane zones in all tissues tested [55]. In developing mouse embryos, transient but preferential expression of collagen XIX was detected in the myotome and coincides with the myogenic transcription factor *myf-5* [56]. In addition, up-regulation of collagen XIX in rhabdomyosarcoma cells has been correlated with myogenic differentiation [57], raising the possibility that collagen XIX is involved in initial stages of skeletal muscle cell differentiation. Very recently, collagen XIX null mice provided unexpected insights regarding the specific role of collagen XIX in morphogenesis. Null mice showed perinatal lethality and the analysis of their phenotype revealed that collagen XIX is required for esophageal muscle transdifferentiation [58]. This finding is in good concordance with the previously established expression pattern of collagen XIX in developing embryos. Notably, collagen XIX expression was confined to few sites including the smooth muscle layers of the stomach and esophagus [56].

2.2.3.3
Collagens XX, XXI and XXII

Chick collagen XX is very closely related to collagen XIV and found to be localized to corneal epithelium, a pattern very close to that of collagen XII [59]. In contrast, collagen XIV was found in corneal stroma revealing a role in corneal compaction [60].

Phylogenic analysis showed that collagen XXI is closely related to collagens XII, XIV and XX [61]. The expression of collagen XXI is developmentally regulated, exhibiting higher level in fetal stages [62]. It localized in several tissues including heart, stomach, kidney, muscle and placenta [63]. More specifically, collagen XXI is an abundant component of the extracellular matrix of vascular walls and is produced by smooth muscle cells. Its expression is stimulated by platelet-derived growth factor, indicating that collagen XXI might contribute to matrix assembly of vascular network during blood vessel formation [62].

The very recently reported collagen XXII was presented as a novel marker of tissue junctions because of its unique tissue distribution. Collagen XXII immunodetection is restricted to the myotendinous and articular cartilage-synovial junctions and to the confined zone between the anagen hair follicle and the dermis [5]. Like other novel FACIT members, this collagen is believed to interact with components of microfibrils rather than with classical collagen fibrils. Tissue junction formation has been poorly investigated because of a lack of a good marker of this specialized region. It is likely that this newly identified collagen will represent an excellent marker to elucidate mechanisms of junction formation during development and regeneration.

2.3 Membrane Collagens and Collagen-Like Proteins

2.3.1 Common Features

The structure and functions of the currently known collagenous transmembrane proteins have been recently reviewed, with a particular emphasis on collagen XVII as a 'prototype' of the group [64]. We will thus focus on characteristic features of this intriguing collagen subfamily.

Membrane collagens are type II transmembrane proteins with the N-terminus inside the cell. This rare orientation occur only in 5% of membrane proteins. Membrane collagens contain a single pass hydrophobic transmembrane domain and include collagens XIII, XVII, XXIII, XVII, and membrane collagen-like proteins corresponding to ectodysplasin, MARCO (macrophage receptor with collagenous structure) and macrophage scavenger receptors [65]. The folding of the triple helix in the ectodomain of collagens XIII [66] and XVII [67] proceeds from the N- to the C-terminus, in opposite orientation to that of the fibrillar collagens. Membrane collagens XIII, XXIII and XXV contain two separate coiled-coil motifs, which may function as independent oligomerization domains [66, 68, 69]. Similar domains occur in the collagen-like membrane proteins MARCO and ectodysplasin-A1 [68]. A high identity was found across the 20-amino acid C-terminal non collagenous domain of membrane collagens XIII, XXIII and XXV [6].

Collagenous transmembrane proteins function as cell surface receptors and soluble matrix molecules shed from the cell surface remaining trimeric after their release in the extracellular space. The cleavage generally occurs close to the extracellular face of the membrane. Ectodomain cleavage at a furin-like site has been reported for collagens XIII [66], XVII [70, 71], XXIII [6], XXV [7] and ectodysplasin [72]. Shed ectodomains of membrane collagens may have a structural and regulatory role and are available for binding to other components of extracellular matrix, such as fibronectin, nidogen-2 and perlecan as reported for collagen XIII [73] or to receptors such as integrins ($\alpha1\beta1$ for collagen XIII, $\alpha5\beta1$ and $\alpha V\beta1$ integrins for collagen XVII [74, 75]). In the same way, shed ectodysplasin binds to its receptor, the EDAR protein [72]. The ectodomains of collagens XIII [73] and XXIII [6] bind to heparin, which inhibits the shedding of collagen XIII in vitro possibly by masking the arginin-rich cleavage site [73]. The release of the ectodomain of collagen XVII from the cell surface is associated with altered keratinocyte motility in vitro [71]. Shedding may also be involved in diseases since the shed ectodomain of collagen XVII is targeted by auto-antibodies in different blistering skin diseases [76].

2.3.2 Membrane Collagens

2.3.2.1 Collagen XIII

Type XIII collagen is a homotrimeric transmembrane collagen composed of a short intracellular domain, a single membrane-spanning region, and an extracellular ectodomain with three collagenous domains separated by short non-collagenous domains. Most of the type XIII collagen ectodomain appears to be present in triple helical conformation [77]. Alternative splicing of collagen XIII affects not only non collagenous sequences but also collagenous domains COL1 and COL3 [78]. Collagen XIII is a component of epidermal cell-matrix and cell-cell contacts in cultured keratinocytes [79]. It is concentrated in cultured skin fibroblasts and several other human mesenchymal cell lines in the focal adhesions and it is found in a range of integrin-mediated adherens junctions in tissues [80]. It is likely involved in the adhesion of cells to basement membranes [73] and seems to participate in the linkage between muscle fiber and basement membrane, a function impaired by the lack of the cytosolic and transmembrane domains [81]. Abnormal adherence junctions in the heart have been observed in transgenic mice overexpressing mutant type XIII collagen, suggesting that it also has an important role in certain adhesive interactions that are necessary for normal development [82].

2.3.2.2 Collagen XVII

Collagen XVII, also referred to as bullous pemphigoid antigen 180 or BP180, is a keratinocyte transmembrane protein, which attaches the epidermis to the basement membrane in the skin [64, 83]. It is a major component of hemidesmosome, where it exists as a full-length protein and interacts with hemidesmosomal components BP230, plectin and the integrin α6β4 [84]. Its extracellular domain is made of 15 collagenous domains interspersed in 16 non-collagenous domains. Its intracellular domain encompasses one third of the molecule and is longer than that of other membrane collagens [64]. The largest collagenous domain of type XVII collagen, COL15, promotes adhesion of epithelial cells and fibroblasts and is targeted by autoantibodies in bullous pemphigoid diseases [85].

2.3.2.3 Collagen XXIII

Collagen XXIII is a transmembrane collagen identified in metastatic tumor cells. It contains only 540 amino acids and consists of an amino-terminal cytoplasmic domain, a transmembrane region, and three collagenous domains

flanked by short noncollagenous domains [6]. It is expressed in normal human heart and retina and its expression is up-regulated in certain metastatic cancer cells [6]. The function of this collagen is still unknown. It may contribute to cell adhesion since it has one RGD site and three KGD motifs, which are involved in cell adhesion and migration on collagen XVII via integrins [75]. Collagen XXIII is able to bind heparin with low affinity and might interact with heparan sulfate chains either on the cell surface or within the extracellular matrix [6].

2.3.2.4
Collagen XXV/CLAC-P (Collagen-Like Alzheimer Amyloid Plaque Component Precursor)

Collagen XXV is the most intriguing membrane collagen. It has been discovered in a search for no-amyloid β components of plaque amyloid in the brain of patients with Alzheimer's disease. Monoclonal antibodies were raised against senile plaque amyloid. The antigen corresponding to the antibody, which binds amyloid fibrils, was purified and the corresponding cDNA cloned [7]. The antigen was named CLAC and its precursor CLAC/P or collagen XXV. It is a short collagen containing 654 amino acids. It is made of three G-X-Y repeat domains flanked by four non collagenous domains. It is expressed specifically in neurons, but not in astrocytes, microglia or meningeal fibroblasts. It is also expressed, albeit at low level, in heart, testis and eye [7]. As reported for collagen XIII in other cell types [82], collagen XXV may play a role in adherens junction between neurons [7]. Its involvement in amyloid deposits in Alzheimer's disease will be discussed below.

2.3.3
The Membrane Collagen-Like Proteins

Ectodysplasin A, MARCO and macrophage scavenger receptors belong to this group.

2.3.3.1
Ectodysplasin A

The two longest isoforms of ectodysplasin (A1 and A2), which are comprised of 391 and 389 amino acids respectively, are collagenous trimeric membrane proteins with a extracellular domain made of 19 repeats of G-X-Y (with an interruption between repeats 11 and 12) and of a C-terminal tumor necrosis factor-like domain [86]. Ectodysplasin co-localizes with cytoskeletal structures at lateral and apical surfaces of cells [87]. It is a signaling molecule required for epithelial morphogenesis [88, 89]. Mutations in the ectodysplasin gene cause X-linked anhidrotic ectodermal dysplasia, a human genetic disorder of impaired ectodermal appendage development (absence or hypoplasia of hair, teeth and sweat glands [90]).

2.3.3.2
Scavenger Receptors Type A and MARCO

The scavenger receptors type A (SR-AI and SR-AII) and MARCO are classical type scavenger receptors that have collagen-like domains. Class A scavenger receptors are transmembrane glycoproteins that mediate both ligand internalization and cell adhesion and participate in lipid metabolism [91]. They are comprised of a short cytoplasmic domain, a transmembrane domain, an α helical coiled-coil domain, a spacer domain and a triple helical domain [92]. In addition, a scavenger receptor cystein-rich domain is present at the C-terminus of SR-AI. The overall structure of MARCO resembles that of SR-AI but the α helical coiled-coil domain is lacking [92–94]. MARCO is involved in the anti-bacterial host defense.

2.3.3.3
Membrane-Type Collectin from Placenta (Collectin Placenta-1)

This protein contains a coiled-coil region, a collagen-like domain and a carbohydrate recognition domain. However, it differs from other collectins in being a type II membrane protein [95] and resembles type A scavenger receptors where the scavenger receptor cysteine-rich domain is replaced by a carbohydrate recognition domain. Collectin placenta-1, which can bind and phagocytose not only bacteria but also yeast, might play important roles in host defenses that are different from those of soluble collectins in innate immunity [96].

2.4
Multiplexins

2.4.1
Common Features

Collagens XV and XVIII are homotrimers with strong sequence and structural homologies in the C-terminal part of the molecule. They contain multiple triple helical domains (9 for collagen XV and 10 for collagen XVIII) hence their name Multiplexins (Multiple Triple Helix with interruptions). They both have at their N-terminus a thrombospondin-1 domain similar to the N-terminal heparin-binding domain of thrombospondin-1 [97], which is also found in some fibrillar collagens and in FACITs. C-terminal heptad repeats have been found in the oligomerization domain of the multiplexins [69].

Both collagens house functional C-terminal domains, endostatin-XVIII and endostatin-XV (also called restin), which are released from the parent molecule by proteolysis as discussed below and share 61% sequence identity. They are both hybrid molecules carrying glycosaminoglycan chains. Collagen XV and XVIII are proteoglycans bearing respectively chondroitin sulfate and heparan sulfate chains [98, 99]. Both collagens occur in the epithelial and endothelial

basement membrane zones of a wide variety of tissues [100]. However, double knockout mice reveal a lack of major functional compensation between collagens XV and XVIII. Their biological role are essentially separate, that of collagen XV in the muscle and that of collagen XVIII in the eye [101].

2.4.2
Collagen XV

The human α1(XV) collagen chain contains a large amino-terminal non-triple helical domain with a tandem repeat structure located in the C-terminal part of the NC10 domain. The 45-amino acid residue sequence, which is not found in collagen XVIII, is repeated four times and is similar to rat cartilage proteoglycan core protein [102].Type XV collagen occurs widely in the basement membrane zones of tissues [103], but its function is not fully elucidated. Lack of type XV collagen in mice results in mild skeletal myopathy and increases vulnerability to exercise-induced skeletal muscle and cardiac injury. Collagen XV appears to function as a structural component needed to stabilize skeletal muscle cells and microvessels [104]. It has been recently hypothesized that collagen XV might be a tumor suppressor [105].

2.4.3
Collagen XVIII

Unlike collagen XV, collagen XVIII is an heparan sulfate proteoglycan and the first reported collagen with a heparan chain [99]. It contains ten domains of triple-helical collagenous repeats separated by non collagenous domains [106].

Collagen XVIII is localized in perivascular basement membrane zones and one of its physiological roles may be to maintain the structural integrity of basement membranes. Indeed, its C-terminal fragment endostatin is often co-localized with perlecan. Perlecan might serve as an adaptator molecule for endostatin via its C-terminal domain endorepellin, which binds to endostatin [107], and could connect collagen XVIII to the basement membrane scaffold [108]. The role of collagen XVIII in basement membrane assembly is strengthened by the presence within its C-terminal fragment of binding sites for heparan sulfate and laminin-1 [109–111] and by the ability of endostatin to bind other components associated to basement membranes, fibulin-1, fibulin-2 and nidogen-2 [112].

A knockout of the collagen XVIII gene in mouse has shown delayed regression of blood vessels in the vitreous body and consequently impaired outgrowth of the retinal vessels, suggesting that collagen XVIII plays a critical role in the development and function of the eye, specially for normal induction of blood vessel formation. Collagen XVIII has been shown to be important for anchoring vitreal collagen fibrils to the inner limiting membrane [113, 114]. This role is further supported by the finding that mutations in COL18A1 lead to an autosomal recessive disorder called Knobloch syndrome, which is defined by the occurrence of vitreoretinal degeneration with retinal detachment, high

myopia, macular abnormalities and occipital encephalocele. The frameshift mutations are located in different regions of the gene [115, 116]. Collagen XVIII has thus a major role in determining the retinal structure as well as in the closure of the neural tube.

Collagen XVIII seems also to have a physiological role during organogenesis. It participates in epithelial branching morphogenesis particularly in the lung and the kidney [117]. It may be involved in various developmental processes through the Wnt signaling pathway via its N-terminal frizzled domain [118]. Indeed the longest N-terminal collagen XVIII variant contains a cysteine-rich region homologous to the extracellular domain of frizzled receptors, which bind the Wnt signaling molecules. In contrast, the C-terminal fragment of the collagen XVIIII molecule, endostatin, has been reported to be a potential inhibitor of Wnt signaling [119]. The structure and functions of endostatin will be detailed in section 3.6. Studies performed in *Caenorhabditis elegans* have shown that collagen XVIII is associated with axons and enriched near synaptic contacts and participates in synapse organization via its C-terminal NC1 domain [120] as discussed below.

2.5 Collagen XXVI and the Emu Family

The members of the Emu (Emilin and Multimerin) family share an N-terminal cysteine-rich domain (EMI domain). Four of these proteins, emilin1, emilin2, emilin and multimerin-domain containing proteins emu1 and emu2, which are secreted and deposited in the extracellular matrix, contain at least a collagenous domain [121].

2.5.1 Emilin1 and Emilin2

Emilins (Elastin Microfibril Interface Located proteINs) are extracellular matrix glycoproteins that are associated with elastic fibers at the interface between amorphous elastin and microfibrils [122]. Both emilins possess a signal peptide, a N-terminal cysteine-rich domain, a long coiled-coil region and a short collagenous sequence followed by a C1q globular domain [122]. The collagenous domain of emilin1 is comprised of 17 uninterrupted G-X-Y triplets and is able to form a triple helix. In emilin2 the 17 triplets have four interruptions and it is unlikely that this sequence could form a triple helix.

2.5.2 Emilin and Multimerin-Domain Containing Proteins (Emu1 and Emu2)

These proteins show a similar structural organization comprised of an N-terminal signal peptide followed by a EMI domain, two collagen stretches and a conserved C-terminal domain of unknown function. These glycosylated proteins

are capable of forming homo- and heterotrimers via disulfide bonding. Alternative splicing gives rise to two isoforms, differing by 2 amino acids, of emu1 and emu2 [121]. It has been suggested that emu1 and emu2 might be incorporated into collagen fibers as macromolecular connectors, similar to FACITs, with the EMI and the C-terminal domains available for further interactions [121].

2.5.3
Collagen XXVI is Identical to the Emu2 Protein

The shortest variant of murine emu2 is 438 amino acid long and contains two collagenous domains corresponding to 23 and 11 G-X-Y repeats respectively. This protein is identical to the α1(XXVI) collagen chain discovered independently by yeast two-hybrid screening of a 17.5 whole mouse embryo cDNA library using the collagen-specific molecular chaperone HSP47 as a bait [8]. Collagen XXVI is expressed in testis and ovary during the development and also in adults. It has not been assigned so far to a collagen group due to a lack of extensive structural similarities with existing sub-families. It is comprised of two collagenous domains and thus is not a fibrillar collagen. It does not appear to be a FACIT because it does not contain either the conserved motif characterized by the presence of two cysteine residues at the COL1-NC1 junction found in other FACITs or a thrombospondin N-terminal-like domain. In contrast, collagen XXVI possesses two coiled-coil oligomerization domains found in membrane collagens XIII, XXIII and XXV, one of them being homologous to that of the membrane collagen XIII [68] but it does not contain a transmembrane sequence and its ectodomain is not cleaved by furin convertase despite the presence of a consensus furin protease cleavage site [8]. Furthermore, this protein has a unique feature relative to other collagens. Whereas in fibrillar collagens, the three α chains are associated in the C-propeptide region where they form intermolecular disulfide bonds, the disulfide bonds that form trimers of collagen XXVI are located in an N-terminal non collagenous domain [8].

2.6
Collagen-Like Molecules

Type II membrane collagen-like molecules have been discussed above. This section deals with molecules containing a collagenous domain, but which are not integral membrane proteins. Although some of the molecules listed below are sometimes referred to as collagens, they are not assigned a Roman number. A letter has been used for the collagenous domain (collagen Q) of acetylcholinesterase [123].

Within this group, a number of proteins contains both a sequence comprised of G-X-Y repeats and a C-terminal globular domain similar to that found in the complement component C1q (see [10–12] for reviews). They include C1q [124], chipmunk hibernation-related proteins, an inner ear-specific structural protein also referred to as saccular collagen [10], the adipocyte-derived hormone

adiponectin [125], C1q-related factor which is not secreted but expressed exclusively in cytoplasm of cells [126], and a novel short-chain collagen called complement C1q tumor necrosis factor-related protein 5 (CQT5) [127]. Mutation in the C1q domain of this protein results in late-onset retinal degeneration, an autosomal dominant disorder with striking clinical and pathological similarities to age-related macular degeneration [127].

Other collagen-like molecules are collectins and ficolins, the humoral lectins of the innate immune defense. They possess a collagenous region and a carbohydrate recognition domain, which is a C-type lectin domain in collectins and a fibrinogen-like domain in ficolins [95]. There are four groups of collectins: the mannan-binding protein group, the surfactant protein A group, the surfactant protein D group, collectin liver 1 and a membrane-type collectin, collectin placenta 1, that functions as a scavenger receptor (see above). The human ficolins are L-ficolin, M-ficolin and H-ficolin. These molecules are sometimes referred to as defense collagens because they bind complex glycoconjugates on microorganisms, thereby inhibiting infection, enhancing the clearance by phagocytes and modulating the immune response [65]. Soluble defense collagens include the complement C1q, collectins, ficolins and adiponectin. Membrane-bound defense collagens are the type A macrophage scavenger receptors and MARCO [65], which were described earlier.

Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins have been recently performed [128]. It is worth noting that bacterial proteins containing collagen-related structural motifs, which are usually found as cell surface or spore components, can form collagen-like triple helices despite the inability of bacteria to synthesize hydroxyproline. Threonine in the Y position of the G-X-Y repeat can substitute for hydroxyproline in stabilizing the triple helix [129].

3 Non Collagenous Domains of Collagens

Comprehensive reviews on the structure of extracellular modules including fibronectin type III and von Willebrand factor type A domains [1] and on their organization in extracellular proteins have been published [2]. The functional role of A-domains in collagen VI has been previously reviewed [130]. We will deal in this section with the structure of individualized non triple-helical domains. We will mostly focus on the possible functions of the Cystein-Rich Repeat and of thrombospondin-1 domains and on the molecular structure of the NC1 domain of collagens IV, VIII, X, XV and XVIII that have been elucidated in the last years. These structures gave insights into the structural basis for collagen molecular and supramolecular assembly and into their biological functions. In addition to functions attributed to individual domains, new functional roles have been reported for a number of proteolytic fragments derived from the C-terminal domain of collagens IV, VIII, XV and XVIII found within,

or in the vicinity of, basement membranes. These enzymatic fragments contain exposed matricryptic sites and are collectively referred to as matricryptins [131]. The matricryptins issued from collagens IV, VIII, XV and XVIII inhibit angiogenesis and tumor growth [132–135]. Among these molecules, endostatin and tumstatin issued from the C-terminal part of α1(XVIII) and α3(IV) chains respectively, are currently extensively investigated.

3.1 Cystein-Rich Repeat Domain

CRRs are found in a number of collagens and other proteins and are characterized by a signature sequence of ten cysteine residues in their sequence (see [136] for review). Physical studies indicated that all ten cysteine residues are disulfide-bonded likely conferring to the domain its stability [137, 138]. This domain is present in multiple copies in two homologous proteins the *Xenopus* chordin and the drosophila sog which bind to members of the TGF-β superfamily, BMP-4 and decapentataplegic respectively. The release of these growth factors by xolloid metalloproteinase activity establishes a gradient of available morphogens that play a crucial role in dorsal-ventral patterning [139, 140]. Particularly interesting has been the speculation of a functional connection between CRR domains of chordin and collagen IIA. Specific binding of BMP-2 and TGF-β1 to recombinant trimeric collagen IIA N-propeptide was clearly established and strongly suggested a role of the collagen IIA N-propeptide in the regulation of growth factor delivery during chondrogenesis [141]. This finding was supported by the recent observation that the *Xenopus* collagen IIA N-propeptide also bound BMP-4 [142]. However, it is too early to speculate that all CRR-containing fibrillar collagens may have similar role in developmental processes. As a result, mice that harbored a deletion of the *col1a1* exon 2 encoding the collagen I CRR-containing domain developed quite normally [143]. Whether collagen III N-propeptide compensated the lack of collagen I NC3 domain or CRR repeats were necessary to bind growth factors efficiently has to be clarified. It has been shown that individual chordin CRRs are less effective than the four repeats present in the intact molecule. Unlike the homotrimeric collagen IIB, collagen I contain only two CRR repeats since the third chain present in the hetrotrimeric molecule α2(I) lacks this domain. Beyond the role of CRR domains in collagens, this finding also questions the exact role of the N-propeptide of collagen I, which was previously implicated in a number of different collagen biogenesis events (molecule assembly, feedback regulation of procollagen synthesis, fibrillogenesis) recently reviewed [136].

3.2 Thrombospondin-1 Domain

The TSPN domain (~200 residues) first described in thrompobondin-1 is present in some fibrillar collagens, in all members of the so-called FACITs sub-

class of collagens and in the multiplexins (see above). The thrombospondin motif occurring in collagens corresponds to the N-terminal heparin-binding domain of thrombospondin, but the amino acid residues involved in heparin binding are not conserved in the collagens.

As described earlier in fibrillar collagens the NC3 domain of the TSPN-containing chains also possesses a variable region that substantially diverges in size and in primary sequence from one chain to another. At the time of this writing, the three dimensional structure of the NC3 domain has not yet been solved. However, rotary shadowing of recombinant of α1(XI) N-terminal domain showed distinct structural organization. The TSPN domain formed a compact globule, whereas the variable region formed an extended tail [144]. The TSPN domain was shown to be released during N-terminal maturation of collagens V and XI but its function as a free released polypeptide or when attached to the procollagens has not been elucidated yet. In point of fact, very little is known about the structure and the function of this domain. A spontaneous mutation in human COL5A1 gene resulting in the deletion of exon 5 encoding the sequence encompassing the BMP-1 cleavage of the N-propeptide caused a connective-tissue disorder, the classical Elhers-Danlos syndrome [145]. Such deletion might abolish the release of TSPN that normally occurred in collagen V maturation and consequently disturbed normal fibrillogenesis. This observation might represent a novel interesting lead to tackle TSPN function.

3.3
Kunitz Domain of the α3(VI) Chain

This domain is located at the C-terminus of the α3(VI) chain. It shows 40–50% identity to a variety of serine protease inhibitors of the Kunitz-type [146], but lacks any inhibitory activity against trypsin, thrombin, kallikrein and several other proteases. Its crystal structure has been determined [146–148] and is very close to that of all other members of the Kunitz family. Two α helices border a twisted antiparallel double-stranded β-structure [146]. The complete domain was shown by NMR to be a highly dynamic molecule in solution, some 44% of its main body structure showing multiple conformations [149].

3.4
NC1 Domains of Collagen IV

Triple-helical collagen IV molecules associate through their N- and C-termini forming a three-dimensional network, which provides basement membranes with an anchoring scaffold and mechanical strength. The six chains of collagen assemble into three different triple helical protomers and self-associate as three distinct networks [150]. The specificity of both molecules and network assembly is governed by amino acid sequences of the C-terminal noncollagenous NC1 domain (~230 residues) of each chain [151].

The crystal structure of the [α1(IV]2α2]2 NC1 hexamer of bovine lens capsule has been determined at 2.0 Å resolution. This was the first hexameric structure to be solved and it provided the structural basis for supramolecular assembly of collagen IV. The football-shaped hexamer is made up of two identical trimers, each containing two α1 chains and one α2 chain. The hexamer structure is stabilized by the extensive hydrophobic and hydrophilic interactions at the trimer-trimer interface without a need for disulfide cross-linking [152]. The NC1 monomer folds into a novel tertiary structure with predominantly β-strands. The [α1(IV]2α2]2 trimer is organized through the unique three-dimensional swapping interactions. The 1.9-Å crystal structure of the NC1 domain of human placenta collagen IV shows stabilization via a covalent Met-Lys cross-link [153]. This hexameric NC1 particle is composed of two trimeric caps, which interact through a large planar interface (Fig. 3). Each cap is formed by two α1 fragments and one α2 fragment with a similar previously uncharacterized fold, segmentally arranged around an axial tunnel. Each trimer forms a quite regular, but nonclassical, sixfold propeller. The trimer-trimer interaction is further stabilized by a previously uncharacterized type of covalent cross-link between the side chains of a methionine and a lysine residue of the α1 and α2 chains from opposite trimers [153]. Three methionine-93 side chains from each cap are cross-connected with the three Lys-211 side chains from the opposite caps.

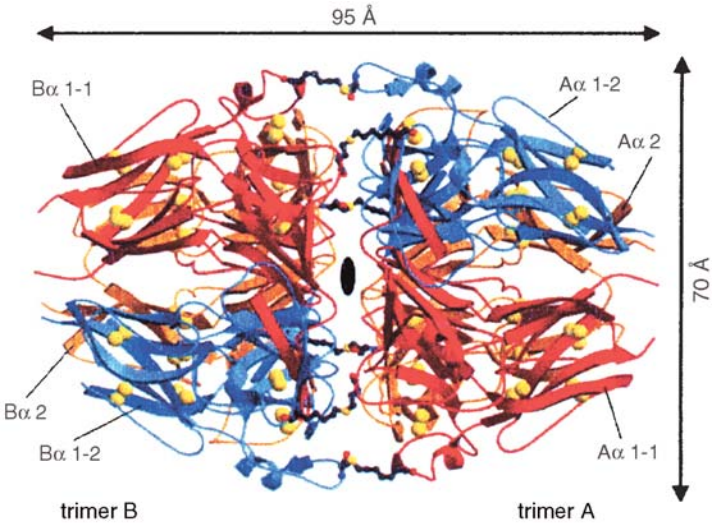

Fig. 3 The human NC1 hexamer of collagen IV. Ribbon plot of the NC1 hexamer viewing along the pseudoexact twofold axis. The two α1(IV) chains and the α2(IV) chain of each trimer are shown in *red* (α111), *blue* (α112) and *gold* (α2). The intrachain disulfides are indicated by their bridging sulfur atoms (*large yellow balls*) and the covalent cross-links between Met-93 and Lys-211 are depicted as *ball-and-stick representations*. Reprinted with permission from [153]. (Copyright 2002 National Academy of Sciences, USA)

The NC1 domain of α1(IV), α2(IV) and α3(IV) chains contains proteolytic fragments called respectively arresten, canstatin and tumstatin, which participate in the control of tumor growth and angiogenesis [133, 134].

3.5 NC1 Domains of Network-Forming Collagens (VIII and X)

Collagens VIII and X are short chain collagens, which share with adiponectin (see above) a common structural organization: a N-terminal domain, a triple helical domain and a C-terminal globular NC1 domain (~160 residues) with a significant sequence homology to the C1q domain of the C1q subunits [154–157]. The determination of the structure of the C1q domain of collagens VIII and X have given insights into the structural basis for their molecular and supramolecular organization in tissues and, in the case of collagen X, in the mechanisms underlying genetic defects (Schmid metaphyseal chondrodysplasia, see below).

The C1q-like C-terminal NC1 domain of collagens VIII and X form stable trimers and are crucial for assembly of collagens VIII and X both at the molecular and supramolecular levels. The driving force for the assembly was suggested to be in the aromatic-rich terminal domain. The C1q domain forms a tightly packed β-sandwich structure related to the tumor necrosis factor fold (Fig. 4 A,B).

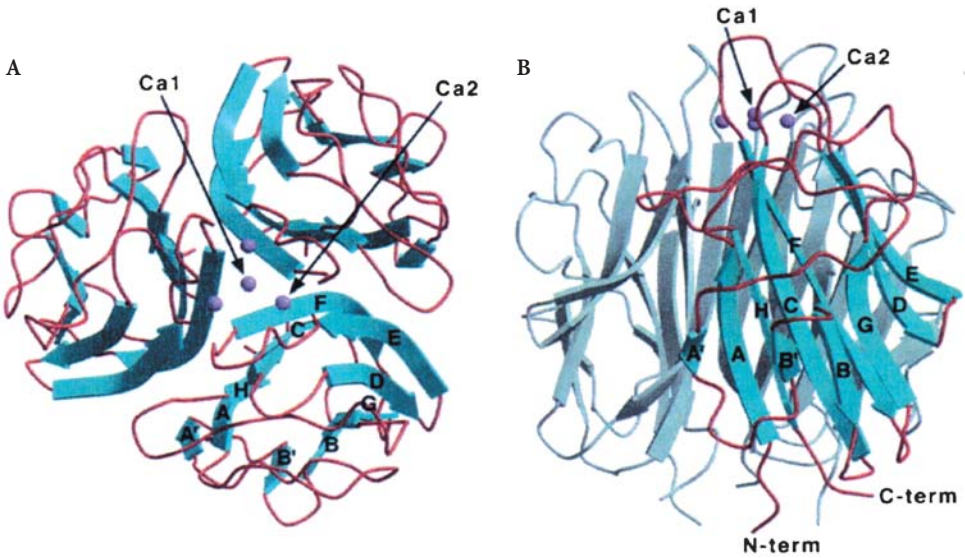

Fig. 4 **A** Structure of the human collagen X NC1 trimer. Cartoon representation of the NC1 trimer viewed down the crystallographic threefold axis. The β strands in one subunit are labeled A; A′, B′ and C-H.Ca^{2+} ions are represented *as pink spheres*. **B** As in A, but rotated by 90° about the horizontal axis. Reprinted from Structure (Canb). 10, Bogin O, Kvansakul M, Rom E, Singer J, Yayon A, Hohenester E. Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen XNC1 domain trimer, 165–173, Copyright (2002), with permission from Elsevier

The overall structures of the NC1 trimer of collagens VIII and X are quite similar. Each subunit of the NC1 domain of collagens VIII and X consists of a ten-stranded β-sandwich with jellyroll topology [158, 159]. The structure of collagen X NC1 trimer, which has been solved at 2.0 Å resolution, reveals an intimate trimeric assembly strengthened by a buried cluster of four calcium ions. The compaction of loops around the Ca^{2+} cluster is likely to contribute significantly to the stability of the trimer. Three strips of exposed aromatic residues on the surface of collagen X NC1 trimer, forming an aromatic patch, are likely to be involved in the supramolecular assembly of collagen X [158]. Equivalent strips exist in the NC1 domain of the closely related collagen VIII, suggesting a conserved assembly mechanism at the supramolecular level [159]. In contrast, the buried calcium cluster present in the collagen X NC1 trimer is not found within the collagen VIII NC1 trimer [159].

The C1q domain of adiponectin is an asymmetrical trimer of β-sandwich protomers, each of which also has a ten-strand jelly-role folding topology identical to TNF-α. The stable trimers formed by the C1q domain are stabilized by a central hydrophobic interface [160].

3.6 Endostatins XVIII and XV

Endostatin is a 20-kDa proteolytic fragment derived from the last 183 amino acids of the non-collagenous C-terminal domain of collagen XVIII and will be further referred to as endostatin-XVIII. It inhibits endothelial cell proliferation and migration, induces apoptosis and inhibits angiogenesis and tumor growth [161, 162]. Endostatin interacts with the $\alpha_5\beta_1$ integrin, and to a lesser extent to $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrins on the surface of endothelial cells. The significance of the integrin binding to endostatin may be to concentrate endostatin at the sites of neovascularization to function as a highly efficient angiogenesis inhibitor by a currently unknown mechanism [163]. Cellular actions and signaling by endostatin are complex issues and β-catenin has been identified as a potential target [164, 165]. Endostatin has entered phase I and II clinical trials in patients with solid tumors (phase I) and neuroendocrine tumors, metastatic and non-amendable for tumor resections (phase II) (see http://cancertrials.nci.nih.gov). Endostatin might also be involved in tumorigenesis in another way, because a coding single nucleotide polymorphism (D104 N) in this fragment predisposes for the development of prostatic adenocarcinoma [166].

The crystal structure of murine [167, 168] and human [169] recombinant endostatin has been solved. Endostatin exhibits a compact fold distantly related to the carbohydrate recognition domain of mammalian C-type lectins and to the hyaluronan link module. Both molecules are centered on a seven-stranded β-sheet and contain loops and 2 α helices, the α1 helix being located on one side of the β-sheet and the α2 helix on the other side [167, 169] (Fig. 5A). Endostatin contains two disulfide bridges in a nested pattern [167]. Structural

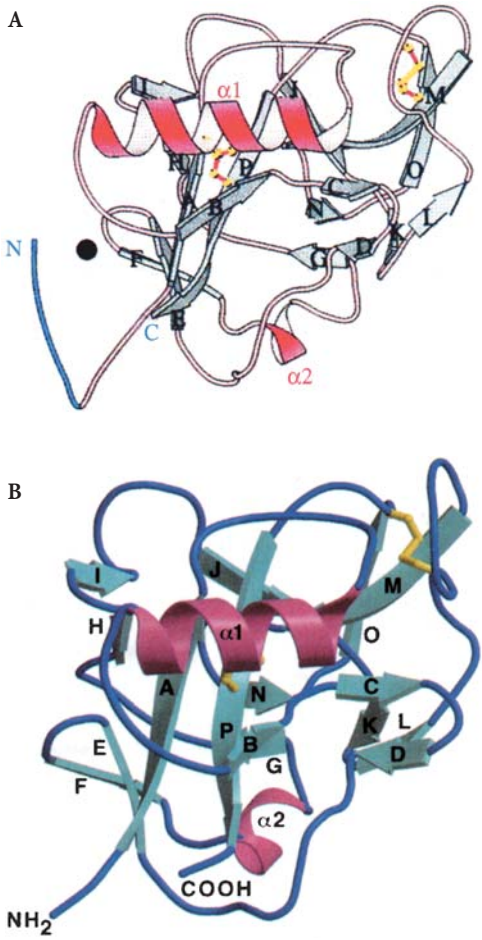

Fig. 5 **A** Structure of human endostatin-XVIII. β-strands (*cyan*) are labeled in sequential order A–P, a helices are *violet* and connecting loops are *pink*. Residues 1–6 are *blue*; zinc is a *black circle*. Reprinted with permission from [169]. **B** Structure of endostatin-XV. Cartoon drawing with β-strands in *light blue*, α-helices in *pink* and disulfide bridges in *yellow*. The polypeptide chain termini are indicated and β-strands are labeled A–P. Reprinted with permission from [112]. (Copyright 1998 National Academy of Sciences, USA)

information has provided useful hints regarding the heparin-binding site of endostatin. Eleven of the 15 positively charged arginine residues cluster on one face of the molecule and, as further investigated by site-directed mutagenesis, this extensive basic patch contains the primary and secondary binding sites to heparin [110]. Human endostatin was found by X-ray crystallography to contain a zinc binding site, which is located at the N-terminus [169]. The coordination varies depending on the crystal [168], but zinc seems to be required for endostatin interaction with heparin/heparan sulfate [170] and for

its anti-angiogenic activity [171] at least in some cases. The modulation of angiogenesis might not be the major physiological role of endostatin [172], which may participate in the anchoring of collagen XVIII to basement membrane as discussed above. Its effects are not restricted to endothelial cells since endostatin has been shown to inhibit the proliferation of tumor cells [170, 173]. Furthermore, the NC1/endostatin domain regulates cell migration and axon guidance in *Caenorhabditis elegans* [174]. The involvement of this domain in neurogenesis is supported by the fact that the deletion of the NC1 domain results in defects in synapse organization and function, suggesting a role for the NC1 domain in the organization of neuromuscular junctions in *Caenorhabditis elegans* [120].

Endostatin-XV, also called restin (for related to endostatin), is the C-terminal proteolytic fragment of collagen XV. Its overall fold [112] is very similar to that of endostatin-XVIII [167]. Like endostatin-XVIII, it contains 2 α helices, 16 β strands and 2 disulfide bridges [112] (Fig. 5B). The face of the molecule that binds heparin in endostatin-XVIII is considerably less basic in endostatin-XV, explaining why endostatin-XV does not bind heparin [112]. In contrast to endostatin-XVIII, endostatin-XV does not bind zinc. Thus, despite a high sequence identity, endostatins derived from collagens XV and XVIII differ in structural and binding properties, tissue distribution and anti-angiogenic activity [112]. Endostatin-XV inhibits bovine aortic endothelial cell migration and proliferation and causes cell apoptosis [175, 176].

4
Collagen-Related Diseases and Prospects for Gene Therapy

Collagen-related diseases comprise genetic and acquired disorders. For comprehensive reviews on autoimmune diseases associated with collagens, the reader is referred to reviews on the pathogenesis of Goodpasture's syndrome [150, 177] and on the subepidermal blistering diseases associated with an immune response to collagens VII [178] and XVII [83, 179]. Collagen genes and their corresponding disorders have been recently reviewed [11, 12, 180] and we will not discuss all of them.

The following section focuses on recently reported mutations (collagens VIII and XVIII) and on genetic diseases for which results concerning gene or cell therapies have been reported because these approaches hold considerable promising for heritable diseases where no effective treatment is currently available. Another field of intense investigation for gene therapy concerns the anti-angiogenic treatment of cancer with endostatin. Several studies have looked into the different possible ways in which endostatin may be administered including gene therapy and cell encapsulation therapy [181–183].

4.1 Osteogenesis Imperfecta

Osteogenesis imperfecta is a varied group of genetic disorders that lead to diminished integrity of connective tissues as a result of alterations in the genes that encode for either the proα1 or proα2 chain of collagen I. Attempts at gene and cell therapies have been performed using marrow stromal cells [184]. Current investigational strategies to treat osteogenesis imperfecta include approaches for skeletal gene therapy and in particular the combined use of genetic manipulation and cellular transplantation. Somatic cell and gene therapies have been evaluated in laboratory and animal studies [185–187].

4.2 Bethlem Myopathy

Collagen VI is an ubiquitous extracellular matrix protein that forms beaded filaments in tissues [10]. Inherited mutations in genes encoding collagen VI in humans cause two muscle diseases, Bethlem myopathy characterized by muscular dystrophy and joint contracture, and Ullrich congenital muscular dystrophy [188]. An unexpected collagen-mitochondria connection [189] has been demonstrated in collagen VI-deficient (Col6a1–/–) mice, which have a muscle phenotype that strongly resembles Bethlem myopathy [190]. The loss of contractile strength is associated with ultrastructural alterations of sarcoplasmic reticulum and mitochondria and spontaneous apoptosis in muscle [190]. The mechanism leading to mitochondrial defects and apoptosis might involve integrins. Lack of collagen VI might cause mitochondrial dysfunction and increased permeability transition pore opening through an abnormal engagement of integrins, in keeping with the fact that integrin-mediated signaling regulates mitochondrial function. Collagen VI myopathies have thus an unexpected mitochondrial pathogenesis that could lead to therapeutic advances [190].

4.3 Alport Syndrome

Collagen IV is a major structural component of basement membranes. In the glomerular basement membrane of the kidney, the α3, α4, and α5(IV) collagen chains form a distinct network that is absent in most patients affected with Alport syndrome, a progressive inherited nephropathy associated with mutation in COL4A3, COL4A4, or COL4A5 genes [150]. Gene therapy of Alport syndrome aims at the transfer of a corrected type IV collagen alpha chain gene into renal glomerular cells responsible for production of the basement membrane. Several studies have been performed in experimental models. Adenovirus-mediated transfer of α5(IV) chain cDNA into swine kidney in vivo leads to the deposition of the protein into the glomerular basement membrane. This indicates that

correction of the molecular defect in Alport syndrome is possible using in vivo gene transfer into glomerular cells [151, 191]. In a canine model of Alport syndrome, transfer of the α5(IV) chain gene to smooth muscle restores in vivo expression of the α6(IV) chain that requires the presence of the α5(IV) chain for incorporation into collagen trimers. The adenoviral vector containing the α(IV) transgene will serve as a useful tool to further explore gene therapy for Alport syndrome [192]. A human-mouse chimera of the α3α4α5(IV) collagen protomer restores a functional glomerular basement membrane and rescues the renal phenotype in Col4a3–/– Alport mice [193].

4.4 Epidermolysis Bullosa

Epidermolysis bullosa comprises a family of inherited blistering skin diseases with variable clinical phenotypes caused by mutations in the COL7A1 gene [178, 194] and in the COL17A1 gene [64, 83]. Mutations in COL17A1 gene are responsible for certain forms of junctional epidermolysis bullosa and a rare subform of epidermolysis bullosa simplex [83], while mutations in the collagen VII gene lead to dystrophic epidermolysis bullosa [178].

Collagen VII is the major constituent of anchoring fibrils, which extend from the lamina densa of epidermal basement membrane into the underlying dermal connective tissue. Because retroviral vectors cannot accommodate the full length collagen VII cDNA, a recombinant truncated type VII collagen minigene has been developed and characterized for gene therapy of dystrophic epidermolysis bullosa. The minigene product retains the functions and characteristics of a full length α_1(VII) chain and its expression in keratinocytes from patients with recessive dystrophic epidermolysis bullosa induced reversion of the recessive dystrophic epidermolysis bullosa phenotype [195]. The suitability of retroviral vectors for gene therapy of this disease has been further demonstrated in dogs expressing a mutated collagen VII [196]. High transfer efficiency, but also high expression levels, are required to ensure therapeutic efficacy in the presence of mutated gene products [196].

Restoration of type VII collagen expression and function in dystrophic epidermolysis bullosa skin in vivo has been successfully performed. The COL7A1 gene was delivered using lentiviral vectors to keratinocytes and fibroblasts from patients with recessive dystrophic epidermolysis bullosa (RDEB). The gene-corrected cells were then used to regenerate human skin on immune-deficient mice [197]. Stable nonviral genetic correction of inherited human skin disease have been reported by introducing COL7A1 gene with a bacteriophage integrase into primary epidermal progenitor cells from patients with RDEB [198]. This circumvents difficulties in stably integrating large corrective sequences such as the large COL7A1 gene into the genomes of long-lived progenitor-cell population. Skin regenerated using these cells displayed stable correction of hallmark RDEB disease features, including collagen VII expression, anchoring fibril formation and dermal-epidermal cohesion [198]. The proof of

principle for genomic DNA vectors (P1-derived artificial chromosome) as a mean of restoring collagen VII production in REB keratinocyte cell line has been reported [199]. Intradermal injection of recessive dystrophic epidermolysis bullosa fibroblasts overexpressing collagen VII into intact RDEB skin stably restored collagen VII expression in vivo and corrected the major features of the disease. Injection of genetically engineered fibroblasts provides a simplified approach to correct human disorders of secreted matrix proteins [200].

Most patients suffering from junctional epidermolysis bullosa lack type XVII collagen mRNA due to nonsense-mediated mRNA decay. A retroviral expression vector for wild-type human collagen XVII has been delivered to primary keratinocytes lacking collagen XVII from patients with junctional epidermolysis bullosa. Restoration of full-length collagen XVII protein expression was associated with adhesion parameter normalization of primary junctional epidermolysis bullosa keratinocytes in vitro [201]. Other approaches and the issues to resolve in this particular field of gene therapy have been reviewed [83, 202]. Future perspectives include the modulation of splicing reactions in keratinocytes, which could be applied in patients with mutations disrupting the normal sequence of a splice site [202].

4.5 Corneal Endothelial Dystrophies

Collagen VIII is a major component of Descemet's membrane, a specialized basement membrane that separates endothelial cells from the stroma in cornea. The first description of mutations in collagen VIII in association with human disease has been reported only in 2001 [203], although this collagen was first described in 1980. Missense mutations in COL8A2, the gene encoding the α2(VIII) chain, cause two forms of corneal endothelial dystrophy.

4.6 Schmid Metaphyseal Chondrodysplasia

Collagen X is involved in chondrocyte hypertrophy and endochondral ossification. Mutations in the COL10A1 gene disrupt growth plate function and result in Schmid metaphyseal chondrodysplasia with short stature and bowed bones. With the exception of two mutations that impair signal peptide cleavage during α1(X) chain biosynthesis, the mutations are clustered within the carboxyl-terminal NC1 domain [10, 204]. However, despite the different nature of the NC1 and signal peptide mutations in collagen X, they all result in impaired collagen X secretion [205].

5
Collagens and Neurodegenerative Diseases

Cultured primary neurons expressed collagen XIII, which enhances neurite outgrowth [206]. Collagen XXV/CLAC-P is specifically expressed by this cell type [7] and endostatin was found in cortical neurons in Alzheimer's disease brains [207]. The longest variant of collagen XVIII is also expressed in human brain [116]. In addition, several collagen types are found to be associated with cerebral deposits of the amyloid β-peptide (Aβ), which are neuropathological lesions characteristic of Alzheimer disease. Pathophysiological hallmarks of the disease are the formation of senile plaque and blood vessels with amyloid angiopathy. The potential of targeting through molecular therapeutics amyloid beta-protein fibrillogenesis, which causes the initiation and progression of Alzheimer's disease, offers an opportunity to improve the disease.

Aβ deposits have been localized to the vascular basement membrane region of capillaries and two collagens found in basement membrane, collagens IV and XVIII, are associated to them. Collagens IV and XVIII are localized in senile plaques in patients with Alzheimer's disease [208 and references therein, 209]. Collagen XVIII is also associated with vascular amyloid Aβ deposits [209] and its C-terminal fragment, endostatin, has been found in neuronal and paracellular deposits in brains of patients with Alzheimer's disease [207]. Endostatin has the propensity to form cross-β-structure and to aggregate into amyloid deposits [210, 211]. Amyloid fibrils formed by endostatin bind and are cytotoxic to murine neuroblastoma cells and to endothelial cells, the cytotoxicity being restricted to the aggregated amyloid form of endostatin [211]. Furthermore, amyloid endostatin induces plasminogen activation by endothelial cells, resulting in vitronectin degradation and plasmin-dependent endothelial cell detachment [212]. This suggests that plasminogen activation system plays a role in endostatin function. Amyloid endostatin may inhibit angiogenesis and tumor growth by stimulating the fibrinolytic system [212].

The extracellular CLAC domain of membrane collagen XXV (see above), generated by furin convertase, is massively deposited within extracellular β amyloid plaque. Both secreted and membrane-tethered forms of collagen type XXV/CLAC-P specifically bind to fibrillized Aβ, implicating these proteins in β-amyloidogenesis and neuronal degeneration in Alzheimer's disease [7]. CLAC is associated with amyloid fibrils that form bundles intermixed with cellular processes and it has been suggested that its three collagenous domains, which are resistant to degradation by a number of proteases, may protect amyloid deposits against degradation [7]. The Alzheimer disease amyloid-associated protein called AMY was found to be identical to CLAC [213].

Another type II transmembrane collagen-like molecule, the scavenger receptor type A, has been shown to bind to the fibrillized form of Aβ and to contribute to the scavenging of Aβ by microglial cells [214]. In Alzheimer's disease, microglial expression of the scavenger receptor type A is increased [215]. The addition of the complement component C1q, which contains a

collagenous domain and is also associated with amyloid deposits, to preformed Aβ aggregates results in significantly increased resistance to aggregate resolubilization [216]. It would be of interest to know if the C1q-related factor, which is expressed at highest levels in the brain [126], can also bind β-amyloid.

In contrast to C1q which promotes fibrillization of Aβ in vitro [216], collagen IV inhibits amyloid β-protein fibril formation in vitro by preventing formation of a β-structured aggregate of Aβ40 and may affect fibril elongation and nucleation [208]. Collagen IV disrupts preformed Aβ42 fibrils in vitro by inducing a structural transition from β-sheet to random structures in Aβ42, which is the initially deposited species in the plaques and can serve as nucleation site for fibril formation by soluble Aβ40 peptide [217].

The ability of collagen IV, and other basement membrane components such as laminin and entactin, to induce disassembly of Aβ fibrils makes them potential effective agents for therapeutics. This also opens new perspectives for the development of novel treatment strategies of Alzheimer's patients to reduce the deposition of the senile plaques in the brain by modulating the amyloidogenic pathway.

6 Conclusion

During the last decade, a number of new collagens have been identified thanks in a large part to the complete sequencing of the human genome. The collagen superfamily should come to completion in a very near future and this may be the opportunity to think about its nomenclature. As a matter of fact some of collagen-like proteins such as emilins fulfill the criteria required to be classified as collagens. Classifying collagen members into separate sub-families is not a simple task, and "cross-talks" between two distinct sub-families are frequent whatever the criteria chosen for their classification: gene structure, sequence homology, structural organization, supramolecular assemblies, and tissue distribution. For instance, the basement membrane collagens includes collagen IV, XV, XVIII and XIX, but if we consider their sequence and structural organization, these collagens fall into three sub-groups: the multiplexins (collagens XV and XVIII), the FACIT collagen XIX and collagen IV, which is the single representative of its group. Another point to consider for future studies is that, starting from collagen XII, all collagens discovered are supposed to form homotrimers. Collagen XXVI might be able to form heterotrimers because Emu2, which is identical to the α1(XXVI) chain as discussed above, is capable of forming heterotrimeric complexes with the Emu1 protein via the EMI domains [121]. It remains to be determined if Emu1 protein can be considered as another α chain of collagen XXVI. Based on their high structural identities, some collagens could form hybrid molecules, notably collagen XXIV and XXVII, as already shown for collagen V and XI. This could dramatically increase the molecular diversity encountered in tissues.

Faced with the remarkable advances in collagen discoveries, the elucidation of their specific properties and functions represents a real challenge for the next future. Collagens do sustain both biomechanical and regulatory functions in almost all tissues. Some of the most striking newest examples have been reviewed in this issue, notably the transmembrane collagens with the neuronal collagen XXV as a component of Alzeimer amyloïd plaques, the identification of the so-called matricryptins including the potent anti-angiogenic endostatin fragment issued from collagen XVIII and the N-propeptide domain, called CRR, of the fibrillar collagen IIA that regulates availability of morphogens during cartilage development.

Most of the recently identified collagens are present in tissues in trace amounts and molecular shape, processing and functional studies were approached by expressing recombinantly the complete triple helical molecule or selected domains. Subsequent to the efficient development of various expression systems, from bacteria to plants (reviewed in [218]), a number of key factors of collagen structure and assembly was elucidated and exciting aspects of unexpected and extremely diverse collagen functions have been revealed. The development of collagen recombinant expression also allows structural studies. The knowledge of three dimensional crystal structure of non collagenous domains and the search for their interacting partners within the extracellular matrix will help to get better understanding of their function. Recently, the co-crystallization of integrin collagen-binding domain with triple helical peptides encompassing the cell binding site revealed the existence of ligand-induced conformational changes, which probably underlie either the affinity regulation or the signaling capacity of integrins [219]. Such methodological development is undoubtedly promising for analyzing cell and/or proteins interaction with collagens. Mice carrying spontaneous or experimentally induced mutations in collagen genes have proven helpful to highlight collagen function during development, to understand collagen diseases and to develop models for human gene therapy [220]. The generation of mutations on the novel collagen genes will likely provide clues on their function. An interesting new development for understanding collagen biology was boosted by the recent complete genome sequencing of *C. elegans*. Some collagen members appeared to be well conserved during evolution [221], the use of this potent developmental organism model will facilitate genetic approaches of collagen functions.

Acknowledgments We apologize to all the investigators whose original work is not quoted in this overview due to space limitations.

References

1. Bork P, Downing AK, Kieffer B, Campbell ID (1996) Q Rev Biophys 29:119
2. Hohenester E, Engel J (2002) Matrix Biol 21:115
3. Christiano AM, Rosenbaum LM, Chung-Honet LC, Parente MG, Woodley DT, Pan TC, Zhang RZ, Chu ML, Burgeson RE, Uitto J (1992) Hum Mol Genet 1:475
4. Hagg P, Rehn M, Huhtala P, Vaisanen T, Tamminen M, Pihlajaniemi T (1998) J Biol Chem 273:15590
5. Koch M, Schulze J, Hansen U, Ashoudt T, Keene DR, Brunker WJ, Burgeson RE, Bruckner P, Bruckner-Tuderman L (2004) J Biol Chem 279:22514
6. Banyard J, Bao L, Zetter BR (2003) J Biol Chem 278:20989
7. Hashimoto T, Wakabayashi T, Watanabe A, Kowa H, Hosoda R, Nakamura A, Kanazawa I, Arai T, Takio K, Mann DM, Iwatsubo T (2002) EMBO J 21:1524
8. Sato K, Yomogida K, Wada T, Yorihuzi T, Nishimune Y, Hosokawa N, Nagata K (2002) J Biol Chem 277:37678
9. Kadler K (1995) Protein Profile 2:491
10. Ricard-Blum S, Dublet B, van der Rest M (2000) Protein Profile Series 2000 (Oxford University Press, ISBN 0198505450)
11. Myllyharju J, Kivirikko KI (2001) Ann Med 33:7
12. Myllyharju J, Kivirikko KI (2004) Trends Genet 20:33
13. Fichard A, Kleman JP, Ruggiero F (1995) Matrix Biol 14:515
14. Hulmes DJ (2002) J Struct Biol 137:2
15. Kielty K, Grant ME (2002) In: Royce PM, Steinman B (eds) Connective tissue and its heritable disorders. Wiley-Liss, New York, pp 159–221
16. Koch M, Laub F, Zhou P, Hahn RA, Tanaka S, Burgeson RE, Gerecke DR, Ramirez F, Gordon MK (2003) J Biol Chem 278:43236
17. Pace JM, Corrado M, Missero C, Byers PH (2003) Matrix Biol 22:3
18. Boot-Handford RP, Tuckwell DS, Plumb DA, Rock CF, Poulsom R (2003) J Biol Chem 278:31067
19. Boot-Handford RP, Tuckwell DS (2003) Bioessays 25:142
20. Exposito JY, Cluzel C, Garrone R, Lethias C (2002) Anat Rec 268:302
21. Fichard A, Tillet E, Delacoux F, Garrone R, Ruggiero F (1997) J Biol Chem 272:30083
22. Chanut-Delalande H, Fichard A, Bernocco S, Garrone R, Hulmes DJ, Ruggiero F (2001) J Biol Chem 276:2435222
23. Sage H, Bornstein P (1979) Biochemistry 18:3815
24. Imamura Y, Scott IC, Greenspan DS (2000) J Biol Chem 275:8749
25. Chernousov MA, Rothblum K, Tyler WA, Stahl RC, Carey DJ (2000) J Biol Chem 275: 28208
26. Chernousov MA, Stahl RC, Carey DJ (2001) J Neurosci 21:6125
27. Ruggiero F, Comte J, Cabanas C, Garrone R (1996) J Cell Sci 109:1865
28. Vogel W (1999) FASEB J13 Suppl:S77
29. Tillet E, Ruggiero F, Nishiyama A, Stallcup WB (1997) J Biol Chem 272:10769
30. Behrendt N, Jensen ON, Engelholm LH, Mortz E, Mann M, Dano K (2000) J Biol Chem 275:1993
31. Delacoux F, Fichard A, Geourjon C, Garrone R, Ruggiero F (1998) J Biol Chem 273:15069
32. Delacoux F, Fichard A, Cogne S, Garrone R, Ruggiero F (2000) J Biol Chem 275:29377
33. Erdman R, Stahl RC, Rothblum K, Chernousov MA, Carey DJ (2002) J Biol Chem 277:7619
34. Eyre DR, Wu JJ, Fernandes RJ, Pietka TA, Weis MA (2002) Biochem Soc Trans 30:893
35. Keene DR, Lunstrum GP, Morris NP, Stoddard DW, Burgeson RE (1991) J Cell Biol 113:971

36. Watt SL, Lunstrum GP, McDonough AM, Keene DR, Burgeson RE, Morris NP (1992) J Biol Chem 267:20093
37. Font B, Aubert-Foucher E, Goldschmidt D, Eichenberger D, van der Rest M (1993) J Biol Chem 268:25015
38. Font B, Eichenberger D, Rosenberg LM, van der Rest M (1996) Matrix Biol 15:341–348
39. Ehnis T, Dieterich W, Bauer M, Kresse H, Schuppan D (1997) J Biol Chem 272:20414
40. Nishiyama T, McDonough AM, Bruns RR, Burgeson RE (1994) J Biol Chem 269: 28193–199
41. Young BB, Gordon MK, Birk DE (2000) Dev Dyn 217:430
42. Fluck M, Giraud MN, Tunc V, Chiquet M (2003) Biochim Biophys Acta 1593:239
43. Myers JC, Li D, Amenta PS, Clark CC, Nagaswami C, Weisel JW (2003) J Biol Chem 278:32047
44. Pan TC, Zhang RZ, Mattei MG, Timpl R, Chu ML (1992) Proc Natl Acad Sci USA 89:6565
45. Yamaguchi N, Kimura S, McBride OW, Hori H, Yamada Y, Kanamori T, Yamakoshi H, Nagai Y (1992) J Biochem (Tokyo) 112:856
46. Yoshioka H, Zhang H, Ramirez F, Mattei MG, Moradi-Ameli M, van der Rest M, Gordon MK (1992) Genomics 13:884
47. Lai CH, Chu ML (1996) Tissue Cell 28:155
48. Kassner A, Hansen U, Miosge N, Reinhardt DP, Aigner T, Bruckner-Tuderman L, Bruckner P, Grassel S (2003) Matrix Biol 22:131
49. Akagi A, Tajima S, Ishibashi A, Yamaguchi N, Nagai Y (1999) J Invest Dermatol 113:246
50. Grassel S, Unsold C, Schacke H, Bruckner-Tuderman L, Bruckner P (1999) Matrix Biol 18:309
51. Grassel S, Timpl R, Tan EM, Chu ML (1996) Eur J Biochem 242:576
52. Tajima S, Akagi A, Tanaka N, Ishibashi A, Kawada A, Yamaguchi N (2000) FEBS Lett 469:1
53. Thein MC, McCormack G, Winter AD, Johnstone IL, Shoemaker CB, Page AP (2003) Dev Dyn 226:523
54. Myers JC, Yang H, D'Ippolito JA, Presente A, Miller MK, Dion AS (1994) J Biol Chem 269:18549
55. Myers JC, Li D, Bageris A, Abraham V, Dion AS, Amenta PS (1997) Am J Pathol 151:1729
56. Sumiyoshi H, Laub F, Yoshioka H, Ramirez F (2001) Dev Dyn 220:155
57. Myers JC, Li D, Rubinstein NA, Clark CC (1999) Exp Cell Res 253:587
58. Sumiyoshi H, Mor N, Lee SY, Doty S, Henderson S, Tanaka S, Yoshioka H, Rattar S, Ramirez F (2004) J Cell Biol 166:591
59. Koch M, Foley JE, Hahn R, Zhou P, Burgeson RE, Gerecke DR, Gordon MK (2001) J Biol Chem 276:23120
60. Gordon MK, Foley JW, Lisenmayer TF, Fitch JM (1996) Dev Dyn 206:49
61. Tuckwell D (2002) Matrix Biol 21:63
62. Chou MY, Li HC (2002) Genomics 79:395
63. Fitzgerald J, Bateman JF (2001) FEBS Lett 505:275
64. Franzke CW, Tasanen K, Schumann H, Bruckner-Tuderman L (2003) Matrix Biol 22:299
65. Tenner AJ (1999) Curr Opin Immunol 11:34
66. Snellman A, Tu H, Vaisanen T, Kvist AP, Huhtala P, Pihlajaniemi T (2000) EMBO J 19:5051
67. Areida SK, Reinhardt DP, Muller PK, Fietzek PP, Kowitz J, Marinkovich MP, Notbohm H (2001) J Biol Chem 276:1594
68. Latvanlehto A, Snellman A, Tu H, Pihlajaniemi T (2003) J Biol Chem 278:37590
69. McAlinden A, Smith TA, Sandell LJ, Ficheux D, Parry DA, Hulmes DJ (2003) J Biol Chem 278:42200
70. Schacke H, Schumann H, Hammami-Hauasli N, Raghunath M, Bruckner-Tuderman L (1998) J Biol Chem 273:25937

71. Franzke CW, Tasanen K, Schacke H, Zhou Z, Tryggvason K, Mauch C, Zigrino P, Sunnarborg S, Lee DC, Fahrenholz F, Bruckner-Tuderman L (2002) EMBO J 21:5026
72. Elomaa O, Pulkkinen K, Hannelius U, Mikkola M, Saarialho-Kere U, Kere J (2001) Hum Mol Genet 10:953
73. Tu H, Sasaki T, Snellman A, Gohring W, Pirila P, Timpl R, Pihlajaniemi T (2002) J Biol Chem 277:23092
74. Nykvist P, Tu H, Ivaska J, Kapyla J, Pihlajaniemi T, Heino J (2000) J Biol Chem 275:8255
75. Nykvist P, Tasanen K, Viitasalo T, Kapyla J, Jokinen J, Bruckner-Tuderman L, Heino J (2001) J Biol Chem 276:38673
76. Schumann H, Baetge J, Tasanen K, Wojnarowska F, Schacke H, Zillikens D, Bruckner-Tuderman L (2000) Am J Pathol 156:685
77. Snellman A, Keranen MR, Hagg PO, Lamberg A, Hiltunen JK, Kivirikko KI, Pihlajaniemi T (2000) J Biol Chem 275:8936
78. Pihlajaniemi T, Rehn M (1995) Prog Nucleic Acid Res Mol Biol 50:225
79. Peltonen S, Hentula M, Hagg P, Yla-Outinen H, Tuukkanen J, Lakkakorpi J, Rehn M, Pihlajaniemi T, Peltonen J (1999) J Invest Dermatol 113:635
80. Hagg P, Vaisanen T, Tuomisto A, Rehn M, Tu H, Huhtala P, Eskelinen S, Pihlajaniemi T (2001) Matrix Biol 19:727
81. Kvist AP, Latvanlehto A, Sund M, Eklund L, Vaisanen T, Hagg P, Sormunen R, Komulainen J, Fassler R, Pihlajaniemi T (2001) Am J Pathol 159:1581
82. Sund M, Ylonen R, Tuomisto A, Sormunen R, Tahkola J, Kvist AP, Kontusaari S, Autio-Harmainen H, Pihlajaniemi T (2001) EMBO J 20:5153
83. Van den Bergh F, Giudice GJ (2003) Adv Dermatol 19:37
84. Koster J, Geerts D, Favre B, Borradori L, Sonnenberg A (2003) J Cell Sci 116:387
85. Tasanen K, Eble JA, Aumailley M, Schumann H, Baetge J, Tu H, Bruckner P, Bruckner-Tuderman L (2000) J Biol Chem 275:3093
86. Wisniewski SA, Kobielak A, Trzeciak WH, Kobielak K (2002) J Appl Genet 43:97
87. Ezer S, Bayes M, Elomaa O, Schlessinger D, Kere J (1999) Hum Mol Genet 8:2079
88. Mikkola ML, Thesleff I (2003) Cytokine Growth Factor Rev 14:211
89. M, Brancaccio A, Weiner L, Missero C, Brissette JL (2003) Genesis 37:30
90. Drogemuller C, Distl O, Leeb T (2003) Genet Sel Evol 35 Suppl 1:S137
91. Platt N, Haworth R, Darley L, Gordon S (2002) Int Rev Cytol 212:1
92. Elomaa O, Sankala M, Pikkarainen T, Bergmann U, Tuuttila A, Raatikainen-Ahokas A, Sariola H, Tryggvason K (1998) J Biol Chem 273:4530
93. Kraal G, van der Laan LJ, Elomaa O, Tryggvason K (2000) Microbes Infect 2:313
94. Sankala M, Brannstrom A, Schulthess T, Bergmann U, Morgunova E, Engel J, Tryggvason K, Pikkarainen T (2002) J Biol Chem 277:33378
95. Holmskov U, Thiel S, Jensenius JC (2003) Annu Rev Immunol 21:547
96. Ohtani K, Suzuki Y, Eda S, Kawai T, Kase T, Keshi H, Sakai Y, Fukuoh A, Sakamoto T, Itabe H, Suzutani T, Ogasawara M, Yoshida I, Wakamiya N (2001) J Biol Chem 276:44222
97. Kivirikko S, Heinamaki P, Rehn M, Honkanen N, Myers JC, Pihlajaniemi T (1994) J Biol Chem 269:4773
98. Li D, Clark CC, Myers JC (2000) J Biol Chem 275:22339
99. Halfter W, Dong S, Schurer B, Cole GJ (1998) J Biol Chem 273:25404
100. Tomono Y, Naito I, Ando K, Yonezawa T, Sado Y, Hirakawa S, Arata J, Okigaki T, Ninomiya Y (2002) Cell Struct Funct 27:9
101. Ylikärppä R, Eklund L, Sormunen R, Muona A, Fukai N, Olsen BR, Pihlajaniemi T (2003) Matrix Biol 22:443
102. Muragaki Y, Abe N, Ninomiya Y, Olsen BR, Ooshima A (1994) J Biol Chem 269:4042
103. Myers JC, Dion AS, Abraham V, Amenta PS (1996) Cell Tissue Res 286:493

104. Eklund L, Piuhola J, Komulainen J, Sormunen R, Ongvarrasopone C, Fassler R, Muona A, Ilves M, Ruskoaho H, Takala TE, Pihlajaniemi T (2001) Proc Natl Acad Sci USA 98:1194
105. Harris H (2003) DNA Cell Biol 22:225
106. Zatterstrom UK, Felbor U, Fukai N, Olsen BR (2000) Cell Struct Funct 25:97
107. Mongiat M, Sweeney SM, San Antonio JD, Fu J, Iozzo RV (2003) J Biol Chem 278:4238
108. Miosge N, Simniok T, Sprysch P, Herken R (2003) J Histochem Cytochem 51:285
109. Sasaki T, Fukai N, Mann K, Gohring W, Olsen BR, Timpl R (1998) EMBO J 17:4249
110. Sasaki T, Larsson H, Kreuger J, Salmivirta M, Claesson-Welsh L, Lindahl U, Hohenester E, Timpl R (1999) EMBO J 18:6240
111. Javaherian K, Park SY, Pickl WF, LaMontagne KR, Sjin RT, Gillies S, Lo KM (2002) J Biol Chem 277:45211
112. Sasaki T, Larsson H, Tisi D, Claesson-Welsh L, Hohenester E, Timpl R (2000) J Mol Biol 301:1179
113. Fukai N, Eklund L, Marneros AG, Oh SP, Keene DR, Tamarkin L, Niemela M, Ilves M, Li E, Pihlajaniemi T, Olsen BR (2002) EMBO J 21:1535
114. Ylikärppä R, Eklund L, Sormunen R, Kontiola AI, Utriainen A, Maatta M, Fukai N, Olsen BR, Pihlajaniemi T (2003) FASEB J 17:2257
115. Sertie AL, Sossi V, Camargo AA, Zatz M, Brahe C, Passos-Bueno MR (2000) Hum Mol Genet 9:2051
116. Suzuki OT, Sertie AL, Der Kaloustian VM, Kok F, Carpenter M, Murray J, Czeizel AE, Kliemann SE, Rosemberg S, Monteiro M, Olsen BR, Passos-Bueno MR (2002) Am J Hum Genet 71:1320
117. Vainio S, Lin Y, Pihlajaniemi T (2003) J Am Soc Nephrol 14 Suppl 1:S3
118. Elamaa H, Snellman A, Rehn M, Autio-Harmainen H, Pihlajaniemi T (2003) Matrix Biol 22:427
119. Hanai J, Gloy J, Karumanchi SA, Kale S, Tang J, Hu G, Chan B, Ramchandran R, Jha V, Sukhatme VP, Sokol S (2002) J Cell Biol 158:529
120. Ackley BD, Kang SH, Crew JR, Suh C, Jin Y, Kramer JM (2003) J Neurosci 23:3577
121. Leimeister C, Steidl C, Schumacher N, Erhard S, Gessler M (2002) Dev Biol 249:204
122. Colombatti A, Doliana R, Bot S, Canton A, Mongiat M, Mungiguerra G, Paron-Cilli S, Spessotto P (2000) Matrix Biol 19:289
123. Massoulie J (2002) Neurosignals 11:130
124. Kishore U, Reid KB (2000) Immunopharmacology 49:159
125. Diez JJ, Iglesias P (2003) Eur J Endocrinol 148:293
126. Berube NG, Swanson XH, Bertram MJ, Kittle JD, Didenko V, Baskin DS, Smith JR, Pereira-Smith OM (1999) Brain Res Mol Brain Res 63:233
127. Hayward C, Shu X, Cideciyan AV, Lennon A, Barran P, Zareparsi S, Sawyer L, Hendry G, Dhillon B, Milam AH, Luthert PJ, Swaroop A, Hastie ND, Jacobson SG, Wright AF (2003) Hum Mol Genet 12:2657
128. Rasmussen M, Jacobsson M, Bjorck L (2003) J Biol Chem 278:32313
129. Xu Y, Keene DR, Bujnicki JM, Hook M, Lukomski S (2002) J Biol Chem 277:27312
130. Shuttleworth A, Ball S, Baldock C, Fakhoury H, Kielty CM (1999) Biochem Soc Trans 27:821
131. Davis GE, Bayless KJ, Davis MJ, Meininger GA (2000) Am J Pathol 156:1489
132. Marneros AG, Olsen BR (2001) Matrix Biol 20:337
133. Kalluri R (2002) Cold Spring Harb Symp Quant Biol 67:255
134. Ortega N, Werb Z (2002) J Cell Sci 115:4201
135. Ricard-Blum S (2003) J Soc Biol 197:41
136. Bornstein P (2002) Matrix Biol 21:217
137. Engel J, Bruckner P, Becker U, Timpl R, Rutschmann B (1977) Biochemistry 16:4026

138. Misenheimer TM, Huwiler KG, Annis DS, Mosher DF (2000) J Biol Chem 275:40938
139. Francois V, Bier E (1995) Cell 80:19
140. Piccolo S, Sasai Y, Lu B, De Robertis EM (1996) Cell 86:589
141. Zhu Y, Oganesian A, Keene DR, Sandell LJ (1999) J Cell Biol 144:1069
142. Larrain J, Bachiller D, Lu B, Agius E, Piccolo S, De Robertis EM (2000) Development 127:821
143. Bornstein P, Walsh V, Tullis J, Stainbrook E, Bateman JF, Hormuzdi SG (2002) J Biol Chem 277:2605
144. Gregory KE, Oxford JT, Chen Y, Gambee JE, Gygi SP, Aebersold R, Neame PJ, Mechling DE, Bachinger HP, Morris NP (2000) J Biol Chem 275:11498
145. Takahara K, Schwarze U, Imamura Y, Hoffman GG, Toriello H, Smith LT, Byers PH, Greenspan DS (2002) Am J Hum Genet 71:451
146. Arnoux B, Merigeau K, Saludjian P, Norris F, Norris K, Bjorn S, Olsen O, Petersen L, Ducruix A (1995) J Mol Biol 246:609
147. Merigeau K, Arnoux B, Perahia D, Norris K, Norris F, Ducruix A (1998) Acta Crystallogr D Biol Crystallogr 54:306
148. Arnoux B, Ducruix A, Prange T (2002) Acta Crystallogr D Biol Crystallogr 58:1252
149. Zweckstetter M, Czisch M, Mayer U, Chu ML, Zinth W, Timpl R, Holak TA (1996) Structure 4:195
150. Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG (2003) N Engl J Med 348:2543
151. Boutaud A, Borza DB, Bondar O, Gunwar S, Netzer KO, Singh N, Ninomiya Y, Sado Y, Noelken ME, Hudson BG (2000) J Biol Chem 275:30716
152. Sundaramoorthy M, Meiyappan M, Todd P, Hudson BG (2002) J Biol Chem 277:31142
153. Than ME, Henrich S, Huber R, Ries A, Mann K, Kuhn K, Timpl R, Bourenkov GP, Bartunik HD, Bode W (2002) Proc Natl Acad Sci USA 99:6607
154. Shuttleworth CA (1997) Int J Biochem Cell Biol 29:1145
155. Sutmuller M, Bruijn JA, de Heer E (1997) Histol Histopathol 12:557
156. Linsenmayer TF, Long F, Nurminskaya M, Chen Q, Schmid TM (1998) Prog Nucleic Acid Res Mol Biol 60:79
157. Plenz GA, Deng MC, Robenek H, Volker W (2003) Atherosclerosis 166:1
158. Bogin O, Kvansakul M, Rom E, Singer J, Yayon A, Hohenester E (2002) Structure (Camb) 10:165
159. Kvansakul M, Bogin O, Hohenester E, Yayon A (2003) Matrix Biol 22:145
160. Shapiro L, Scherer PE (1998) Current Biol 8:335
161. Sasaki T, Hohenester E, Timpl R (2002) IUBMB Life 53:77
162. Ren B, Hoti N, Rabasseda X, Wang YZ, Wu M (2003) Methods Find Exp Clin Pharmacol 25:215
163. Rehn M, Veikkola T, Kukk-Valdre E, Nakamura H, Ilmonen M, Lombardo C, Pihlajaniemi T, Alitalo K, Vuori K (2001) Proc Natl Acad Sci USA 98:1024
164. Ramchandran R, Karumanchi SA, Hanai J, Alper SL, Sukhatme VP (2002) Crit Rev Eukaryot Gene Expr 12:175
165. Dixelius J, Cross MJ, Matsumoto T, Claesson-Welsh L (2003) Cancer Lett 196:1
166. Iughetti P, Suzuki O, Godoi PH, Alves VA, Sertie AL, Zorick T, Soares F, Camargo A, Moreira ES, di Loreto C, Moreira-Filho CA, Simpson A, Oliva G, Passos-Bueno MR (2001) Cancer Res 61:7375
167. Hohenester E, Sasaki T, Olsen BR, Timpl R (1998) EMBO J 17:1656
168. Hohenester E, Sasaki T, Mann K, Timpl R (2000) J Mol Biol 297:1
169. Ding YH, Javaherian K, Lo KM, Chopra R, Boehm T, Lanciotti J, Harris BA, Li Y, Shapiro R, Hohenester E, Timpl R, Folkman J, Wiley DC (1998) Proc Natl Acad Sci USA 95:10443

170. Ricard-Blum S, Feraud O, Lortat-Jacob H, Rencurosi A, Fukai N, Dkhissi F, Vittet D, Imberty A, Olsen BR, van Der Rest M (2004) J Biol Chem 279:2927
171. Boehm T, O'Reilly MS, Keough K, Shiloach J, Shapiro R, Folkman J (1998) Biochem Biophys Res Commun 252:190
172. From the Editor's desk (2002) Matrix Biol 21:309
173. Dkhissi F, Lu H, Soria C, Opolon P, Griscelli F, Liu H, Khattar P, Mishal Z, Perricaudet M, Li H (2003) Hum Gene Ther 14:997
174. Ackley BD, Crew JR, Elamaa H, Pihlajaniemi T, Kuo CJ, Kramer JM (2001) J Cell Biol 19:1219
175. Ramchandran R, Dhanabal M, Volk R, Waterman MJ, Segal M, Lu H, Knebelmann B, Sukhatme VP (1999) Biochem Biophys Res Commun 255:735
176. Xu R, Xin L, Fan Y, Meng HR, Li ZP, Gan RB, Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai) (2002) 34:138
177. Borza DB, Neilson EG, Hudson BG (2003) Semin Nephrol 23:522
178. Bruckner-Tuderman L, Hopfner B, Hammami-Hauasli N (1999) Matrix Biol 18:43
179. Zillikens D (2002) Keio J Med 51:21
180. Byers PH (2000) Clin Genet 58:270
181. Nguyen JT (2000) Adv Exp Med Biol 465:457
182. Sorensen DR, Read TA (2002) Int J Exp Pathol 83:265
183. Visted T, Lund-Johansen M (2003) Expert Opin Biol Ther 3:551
184. Prockop DJ (1999) Biochem Soc Trans 27:15
185. Cole WG (2002) Clin Orthop 401:6
186. Niyibizi C, Wallach CJ, Mi Z, Robbins PD (2002) Crit Rev Eukaryot Gene Expr 12:163
187. Niyibizi C, Wang S, Mi Z, Robbins PD (2004) Gene Ther Jan 15 [Epub ahead of print]
188. Bertini E, Pepe G (2002) Eur J Paediatr Neurol 6:193
189. Rizzuto R (2003) Nat Genet 35:300
190. Irwin WA, Bergamin N, Sabatelli P, Reggiani C, Megighian A, Merlini L, Braghetta P, Columbaro M, Volpin D, Bressan GM, Bernardi P, Bonaldo P (2003) Nat Genet 35: 367
191. Heikkila P, Tibell A, Morita T, Chen Y, Wu G, Sado Y, Ninomiya Y, Pettersson E, Tryggvason K (2001) Gene Ther 8:882
192. Harvey SJ, Zheng K, Jefferson B, Moak P, Sado Y, Naito I, Ninomiya Y, Jacobs R, Thorner PS (2003) Am J Pathol 162:873
193. Heidet L, Borza DB, Jouin M, Sich M, Mattei MG, Sado Y, Hudson BG, Hastie N, Antignac C, Gubler MC (2003) Am J Pathol 163:1633
194. Pulkkinen L, Uitto J (1999) Matrix Biol 18:29
195. Chen M, O'Toole EA, Muellenhoff M, Medina E, Kasahara N, Woodley DT (2000) J Biol Chem 275:24429
196. Baldeschi C, Gache Y, Rattenholl A, Bouille P, Danos O, Ortonne JP, Bruckner-Tuderman L, Meneguzzi G (2003) Hum Mol Genet 12:1897
197. Chen M, Kasahara N, Keene DR, Chan L, Hoeffler WK, Finlay D, Barcova M, Cannon PM, Mazurek C, Woodley DT (2002) Nat Genet 32:670
198. Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q, Fang M, Calos MP, Khavari PA (2002) Nat Med 8:1166 (Erratum in Nat Med 2003 9:237)
199. Mecklenbeck S, Compton SH, Mejia JE, Cervini R, Hovnanian A, Bruckner-Tuderman L, Barrandon Y (2002) Hum Gene Ther 13:1655
200. Ortiz-Urda S, Lin Q, Green CL, Keene DR, Marinkovich MP, Khavari PA (2003) J Clin Invest 111:251
201. Seitz CS, Giudice GJ, Balding SD, Marinkovich MP, Khavari PA (1999) Gene Ther 6:42
202. Bauer JW, Lanschuetzer C (2003) Clin Exp Dermatol 28:53

203. Biswas S, Munier FL, Yardley J, Hart-Holden N, Perveen R, Cousin P, Sutphin JE, Noble B, Batterbury M, Kielty C, Hackett A, Bonshek R, Ridgway A, McLeod D, Sheffield VC, Stone EM, Schorderet DF, Black GC (2001) Hum Mol Genet 10:2415
204. Bateman JF (2001) Osteoarthritis Cartilage Suppl A:S141
205. Chan D, Ho MS, Cheah KS (2001) J Biol Chem 276:7992
206. Sund M, Vaisanen T, Kaukinen S, Ilves M, Tu H, Autio-Harmainen H, Rauvala H, Pihlajaniemi T (2001) Matrix Biol 20:215
207. Deininger MH, Fimmen BA, Thal DR, Schluesener HJ, Meyermann R (2002) J Neurosci 22:10621 (Erratum in J Neurosci 2003 23:725)
208. Kiuchi Y, Isobe Y, Fukushima K (2002) Life Sci 70:1555
209. van Horssen J, Wilhelmus MM, Heljasvaara R, Pihlajaniemi T, Wesseling P, de Waal RM, Verbeek MM (2002) Brain Pathol 12:456
210. Kranenburg O, Bouma B, Kroon-Batenburg LM, Reijerkerk A, Wu YP, Voest EE, Gebbink MF (2002) Curr Biol 12:1833
211. Kranenburg O, Kroon-Batenburg LM, Reijerkerk A, Wu YP, Voest EE, Gebbink MF (2003) FEBS Lett 539:149
212. Reijerkerk A, Mosnier LO, Kranenburg O, Bouma BN, Carmeliet P, Drixler T, Meijers JC, Voest EE, Gebbink MF (2003) Mol Cancer Res 1:561
213. Soderberg L, Zhukareva V, Bogdanovic N, Hashimoto T, Winblad B, Iwatsubo T, Lee VM, Trojanowski JQ, Naslund J (2003) J Neuropathol Exp Neurol 62:1108
214. El Khoury J, Hickman SE, Thomas CA, Cao L, Silverstein SC, Loike JD (1996) Nature 382:716
215. Husemann J, Loike JD, Anankov R, Febbraio M, Silverstein SC (2002) Glia 40:195
216. Webster S, O'Barr S, Rogers J (1994) J Neurosci Res 39:448
217. Kiuchi Y, Isobe Y, Fukushima K, Kimura M (2002) Life Sci 70:2421
218. Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M, Polarek J (2003) Adv Drug Deliv Rev 55:1547
219. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC (2000) Cell 101:47
220. Aszodi A, Pfeifer A, Wendel M, Hiripi L, Fassler R (1998) J Mol Med 76:238
221. Johnstone IL (2000) Trends Genet 16:21

Top Curr Chem (2005) 247: 85–114
DOI 10.1007/b103820
© Springer-Verlag Berlin Heidelberg 2005

Collagen Biosynthesis

Takaki Koide[1] · Kazuhiro Nagata[2] (✉)

[1] Faculty of Pharmaceutical Science, Niigata University of Pharmacy and Applied Life Science, 5-13-2 Kamishin'ei-cho, Niigata 950-2081, Japan
koi@niigata-pharm.ac.jp

[2] Department of Molecular and Cellular Biology, Institute for Frontier Medical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8397, Japan
nagata@frontier.kyoto-u.ac.jp

1 Overview of Procollagen Biosynthesis 86

2 Initial Trimerization of α-Chains 89

3 Nucleation and Propagation of the Triple Helix 92
3.1 Nucleation of the Triple Helix 92
3.2 Propagation of the Triple Helix 93

4 Molecular Chaperones Involved in Procollagen Biosynthesis 94
4.1 Bip/Grp78 96
4.2 Calnexin and Calreticulin 96
4.3 Protein Disulfide Isomerase (PDI) 97
4.4 Prolyl 4-hydroxylase (P4-H) 98
4.5 Peptidyl-Prolyl *cis-trans* Isomerase (PPIase) 99
4.6 Hsp47 100

5 Intracellular Trafficking of Procollagen Molecules 105
5.1 ER-to-Golgi Transport 105
5.2 Intra- and Post-Golgi Transport 106

6 Production of Recombinant Collagens 107

7 Conclusions and Future Perspectives 108

References 110

Abstract Collagen is synthesized in the endoplasmic reticulum (ER) as procollagen, which is the precursor protein that bears propeptide domains at either end of the triple helical domain. The processes by which procollagen is synthesized in the lumen of the ER include unique steps that are not found in the biosynthesis of globular proteins. First, each polypeptide chain of procollagen (proα-chains) finds its correct partners, which enables the formation of the distinct types of procollagen. Second, triple helix-formation of long Gly-X-Y repeats starts at a defined region, which results in the formation of a correctly aligned triple helix and thereby prevents mis-staggering. The most characteristic step is the formation of the triple helix. This step involves specific post-translational modifications, in particular, the prolyl 4-hydroxylation of the Y-position amino acids that stabilizes the triple

helical conformation. The formation of the triple helix is a slow process compared to the folding of globular proteins, including *cis-trans* isomerization of the many prolyl and hydroxyprolyl peptide bonds. Recent advances have indicated that these processes are assisted by a set of the ER-resident molecular chaperones, such as protein disulfide isomerase (PDI), peptidyl prolyl *cis-trans* isomerases (PPIases), heat-shock protein (Hsp)47, and prolyl 4-hydroxylase (P4-H). The intracellular trafficking of procollagen molecules has also been shown to involve a pathway distinct from that utilized by small secretory proteins.

Keywords Procollagen · Triple helix · Molecular chaperone · Folding · Endoplasmic reticulum

List of Abbreviations

BSE	Bovine spongiform encephalopathy
C1q	Complement 1q
collectin	Collagen-like lectin
dpc	Days post coitus
ECM	Extracellular matrix
ER	Endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
ES cell	Embryonic stem cell
FACIT	Fibril-associated collagens with interrupted triple helices
GFP	Green fluorescent protein
Hsp47	47-kDa heat-shock protein
Hyl	Hydroxylysine
Hyp	4-Hydroxyproline
P4-H	Prolyl 4-hydroxylase
PDI	Protein disulfide isomerase
PPIase	Peptidyl-prolyl *cis-trans* isomerase
SERPIN	Serine protease inhibitor
SLS	Segment-long-spacing
UPR	Unfolded protein response

1 Overview of Procollagen Biosynthesis

Collagens occur ubiquitously in multicellular animals as they are the major components of the extracellular matrices (ECMs). The feature that characterizes the collagens is that they consist of tandemly repeated Gly-X-Y triplets and have a unique triple helical structure. The number of triplet repeats ranges from a few dozen to 510 depending on the collagen type. Some types of collagen are homotrimeric proteins that consist of three identical α-chains, such as type II and III, while others are heterotrimers of different α-chains. For example, type I collagen consists of two α1(I) chains and one α2(I) chain.

To date, 27 types of collagen consisting of specific α-chains that are encoded by over 40 different genes have been identified. Most collagen types assemble into supramolecular architectures either by themselves or with the aid of other

ECM components, including other types of collagen. The collagen family proteins are classified into the following subfamilies based on their molecular and supramolecular structures:

1. Fibril-forming collagens (types I, II, III, V and XI)
2. Network-forming collagens (types IV, VIII and X)
3. Fibril-associated collagens with interrupted triple helices (FACITs; types IX, XII, XIV, XVI and XIX)
4. Transmembrane collagens (types XIII and XVII)
5. Other types of collagens

In addition, a number of collagen-related proteins, such as complement 1q (C1q), lung surfactant proteins, collagen-like lectins (collectins) [1, 2], the tail of the asymmetric form of acetylcholinesterase [3], and macrophage scavenger receptors [4, 5], also possess collagenous triple helices composed of Gly-X-Y repeats.

All of the collagen family proteins are secretory proteins that pass through the endoplasmic reticulum (ER), where they are folded into triple helical molecules. All collagens also contain non-collagenous domains; in the case of the fibril-forming collagens; these are recognized as propeptide domains. In this chapter, we focus on the pathway by which collagen is synthesized within the cell. Unless otherwise indicated, this chapter will center on the biosynthesis of the fibril-forming collagens since most of the studies on collagen biosynthesis have examined this type of collagen (mainly types I and III).

The fibril-forming collagens form similar molecular and supramolecular architectures as they all generate cross-striated fibrils with an axial D periodicity that is typically 67 nm. The individual molecules consist of long triple helical domains comprised of approximately 330 Gly-X-Y triplets per chain, and these form rope-like molecules with a length of 300 nm. The biosynthesis of these collagen structures involves a unique pathway that consists of a triple helix-forming process coupled with a number of specific post-translational modifications. Collagen is first synthesized and secreted as procollagen, a larger precursor protein (Fig. 1). These procollagen molecules possess a central triple helical domain and non-collagenous propeptide domains at both their N- and C-terminal ends. Both propeptide domains are cleaved off in the final step of collagen biosynthesis.

Procollagen biosynthesis was first outlined in the late 1970s [6]. The pathway that was described then is basically compatible with the current knowledge in this research field (Fig. 2). Since each proα-chain contains the signal sequence at its N-terminus, the nascent polypeptide is co-translationally translocated into the lumen of the rough ER, like other secretory proteins. Post-translational modifications of the triple helix-forming region are initiated during this translocation step [7]. The post-translational modifications include prolyl 4-hydroxylation at the Y-positions, prolyl 3-hydroxylations at the X-positions, and lysyl hydroxylations to form hydroxylysine (Hyl) residues. Some of the Hyl residues are subsequently glycosylated to form galactosyl-Hyl and glucosyl-

Fig. 1 Structure and shape of procollagen

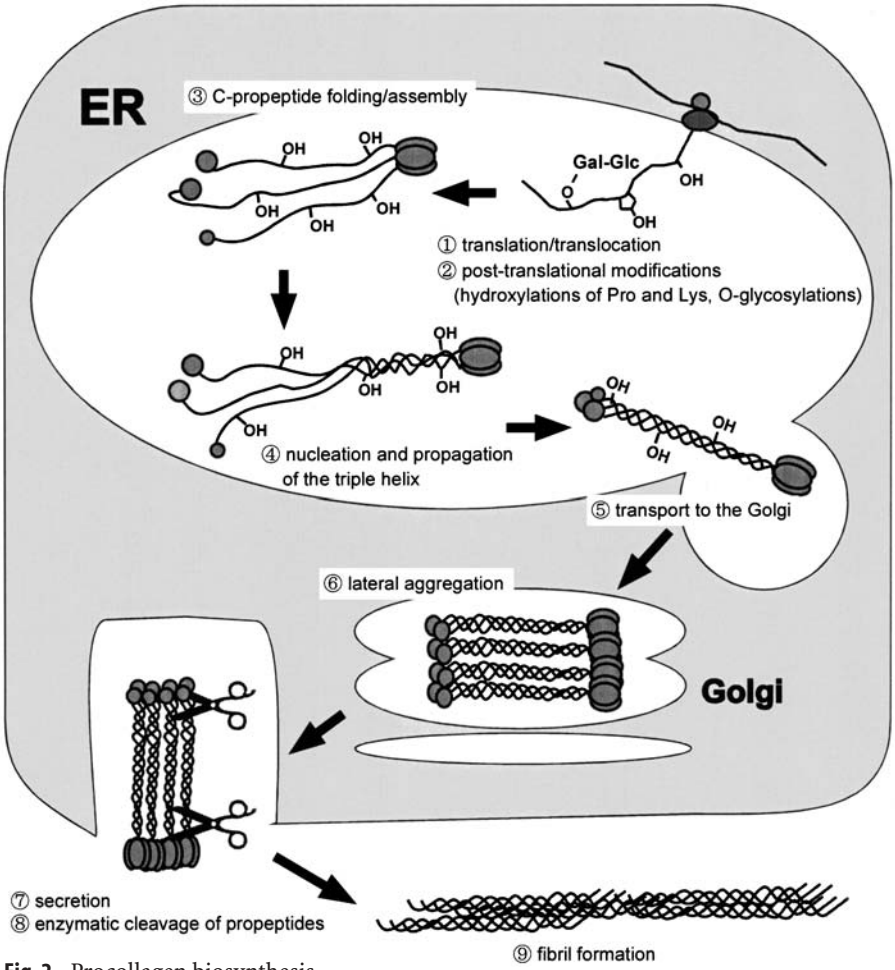

Fig. 2 Procollagen biosynthesis

galactosyl-Hyl residues. After the polypeptide translocation is completed, three proα-chains trimerize at the C-propeptide domains. The association of the C-propeptides is stabilized by intermolecular disulfide bridges. It should be noted that the association of the C-propeptides must be accomplished with correct pairing according to the collagen type.

Following C-propeptide assembly, the formation of the triple helix starts from a C-terminal nucleus and then propagates toward the N-terminus similar to fastening a zipper. This triple helix-forming process involves prolyl 4-hydroxylation of up to 100 Pro residues per chain; these increase the thermal stability of the triple helix. This is shown, for instance, by the fact that the blocking of prolyl 4-hydroxylation by α, α′-dipyridyl, which chelates the Fe(II) ions that are essential for prolyl 4-hydroxylase (P4-H) activity, inhibits intracellular triple helix formation [8].

The procollagen molecules that have completed their modifications and folding are then transported to the Golgi apparatus. In the Golgi cisternae, the procollagen molecules are stacked laterally and form aggregates [9, 10]. Finally, the procollagen aggregates are secreted into the extracellular space, where the N- and C-propeptides are enzymatically cleaved off, thus generating the mature collagen molecules. Upon cleavage of the propeptides, the triple helical collagen molecules self-assemble into fibrillar supramolecules in which each molecule is displaced about one-quarter of its length along the axis of the fibril.

The events that occur in the ER involve successive actions of ER-resident molecular chaperones that play either a general role in chaperoning a variety of secretory proteins or a specialized role for procollagen. These molecular chaperones also contribute to the quality control mechanism that ensures that only correctly-folded procollagen is secreted.

In the following sections of this chapter, we will describe each step of collagen biosynthesis in detail, including recent observations. We will also describe how the various molecular chaperones participate in procollagen biosynthesis.

2 Initial Trimerization of α-Chains

The folding of procollagen begins with C-propeptide trimerization after the complete translocation of the proα-chains into the lumen of the ER. Before triple helix-formation, the trimerized C-propeptide domains are stabilized by intermolecular disulfide bridges [11–14].

Since the arrangement of the α-chains in the trimer varies depending on the types of collagen involved, the α-chains must associate in a distinct and controlled manner that eliminates the possibility of generating hybrid α-chains that are composed of different types of collagens. Exceptions to this are proα1(XI) and proα2(XI), which can form hybrid heterotrimers with other proα-chains that produce type V collagens in certain tissues [15–17]. Most collagen-producing cells simultaneously synthesize different types of procollagen

molecules without producing incorrectly assembled α-chains. For instance, skin fibroblasts produce at least three fibril-forming collagens, namely, types I, III and V, which are comprised of six homologous α-chains, α1(I), α2(I), α1(III), α1(V), α2(V) and α3(V). How can these different but homologous chains assemble in such a selective manner? The triple helices themselves do not possess an intrinsic ability to find the correct partners. Unlike the base pair-formation that results in a DNA duplex, the Gly-X-Y sequences in procollagen do not complementarily recognize one another in a residue-to-residue fashion. Rather, it is the C-propeptide domains of procollagen that assure the type-specific chain assembly of the individual procollagen chains. Observations of fibril-forming procollagen molecules that contain naturally occurring mutations or artificial deletions have revealed the important roles played by the C-propeptide in the trimerization that is the initial step of procollagen folding [18, 19].

The C-propeptides of the proα-chains of fibril-forming collagens are non-collagenous domains containing approximately 250 amino acid residues. These C-propeptides are highly homologous to one other [20] (Fig. 3A–C) as they contain eight conserved cysteine residues that are oxidized into four cystine pairs during the folding/trimerization process. The four cysteine residues located in the N-terminus of the C-propeptide contribute to this intermolecular crosslinking while the latter four form intramolecular disulfide bridges [21, 22]. Although the atomic-level structure of a procollagen C-propeptide trimer has not been solved, a low resolution model of the trimer of the recombinantly expressed C-propeptide of human type III procollagen has been generated by laser light scattering, analytical centrifugation, and small angle X-ray scattering [23, 24]. This revealed the C-propeptide trimer as a cruciform structure with three major lobes and one minor lobe (Fig. 3E). It was speculated that the three major lobes are the individually folded C-terminal regions of the C-propeptides and the minor lobe consists of the disulfide-bridged N-terminal regions of the C-propeptides that are connected to the triple helix region.

Bulleid and co-workers have demonstrated by swapping the C-propeptides of procollagens that the C-propeptide domains are necessary and sufficient to direct the assembly of homotrimeric procollagens with correctly aligned triple helices [25]. The region within the C-propeptide that is responsible for the type-specific proα-chain association has been found in the peptide sequence of 23 amino acid residues. Within this sequence, 15 discontinuous amino acid residues have been shown to be responsible for the chain selectivity [25] (Fig. 3C).

Another region that probably contributes to the associations of the proα-chains was recently found, namely, the α-helical coiled-coil motifs located at the N-termini of the C-propeptide domains [26]. The coiled-coil motif contains heptad repeats (*a-b-c-d-e-f-g*) in which positions *a* and *d* are occupied by hydrophobic amino acid residues [27]. The α-helical coiled-coil structure had been shown to be present in various collectins [2] and other collagen-like proteins, including macrophage scavenger receptors [4, 5]. MacAlinden et al. have recently shown that not only these collagen-like proteins but also most types of procollagen contain 2–4 heptad repeat sequences that should fold into three-

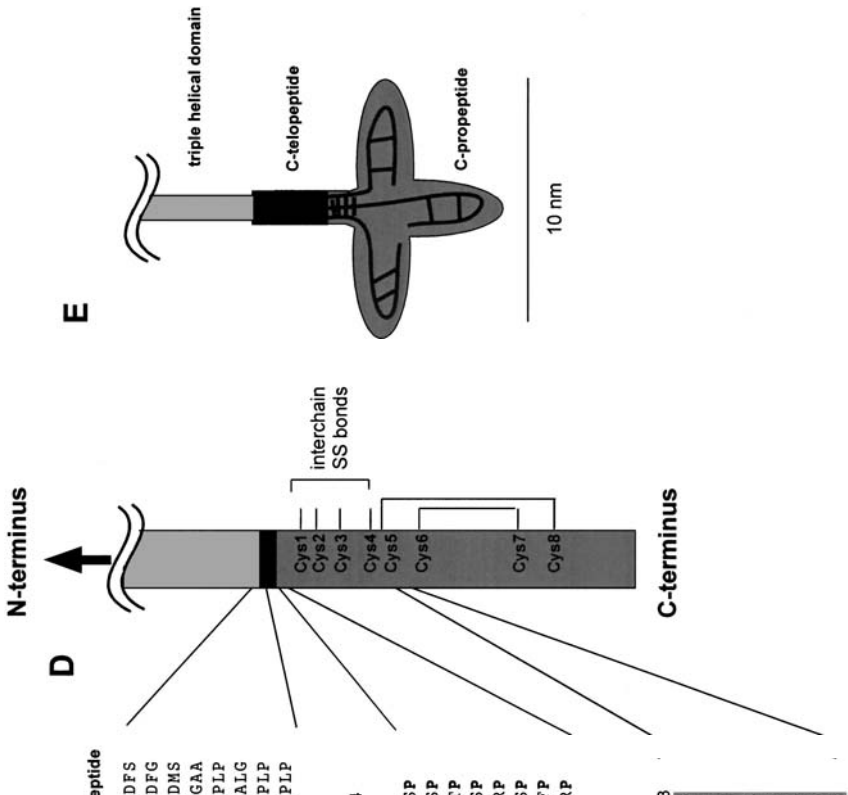

Fig. 3 Structure of C-propeptide
A–D regions responsible for trimerization, **E** proposed model of the C-propeptide domain

stranded coiled-coils (Fig. 3B). Indeed, they have actually shown that the peptide with the heptad repeat sequence of the type II procollagen C-propeptide forms a trimer that has an α-helical coiled-coil structure [26]. It is likely that the C-propeptide coiled-coil region is sufficient to drive the trimerization of procollagen molecules lacking the rest of the C-propeptide because when the rest of the C-propeptide is replaced with the transmembrane domain of the trimeric influenza virus haemagglutinin, it can still form trimers and subsequently generate the triple helix [28]. However, such a coiled-coil formation is only possible when the part of the C-propeptide that contains the molecular recognition sequence of 15 amino acid residues is also present [26].

3
Nucleation and Propagation of the Triple Helix

3.1
Nucleation of the Triple Helix

After the three proα-chains associate at the C-propeptide domains, the trimer is stabilized by forming interchain disulfide bridges. In the case of type III procollagen, additional interchain disulfide bridges are formed in the C-telopeptide region, which is a short non-collagenous sequence that links the triple helical domain with the C-propeptide. The triple-helical regions, which can be as long as 1000 amino acid residues, then start to twine around one another, thereby propagating the formation of the triple helix toward the N-termini.

The collagen triple helix is a right-handed supercoil consisting of three polyproline II-like left-handed helices in which all peptide bonds are in the *trans* configuration [27, 29, 30]. To form the triple helix, it is essential that every third residue is a Gly; no other amino acid can replace these Gly residues. The Gly residues are buried in the core of the helix. Unlike the α-helical coiled-coil, the collagen triple helix does not bear direct intramolecular hydrogen bonds between the main-chain amide and the carbonyl groups; instead, all the NH-CO hydrogen bonds are formed only between the different polypeptide chains. This means the triple helix formation is probably coupled to the concomitant folding of individual proα-chains into the polyproline II-helix. In addition, since the collagen triple helix does not possess a hydrophobic core, the formation of the triple helix is not expected to employ the hydrophobic interactions that most other proteins utilize in their folding.

To form a correctly aligned triple helical procollagen molecule, the initial nucleation of the helix must occur at one distinct site. If triple helix formation is simultaneously initiated at multiple nucleation sites, mis-aligned and hence only partially triple helical molecules like gelatin would be produced. Thus, there must be a mechanism in the procollagen folding pathway that either directs the nucleation at a distinct site and/or prevents unfavorable nucleation of the triple helix. It has been suggested that the C-terminal end of the triple

helix-forming domain nucleates first in order to facilitate C to N zipper-like folding. Bulleid and co-workers have elucidated the mechanism of the nucleation of the triple helix by utilizing the semi-permeabilized cell system that enables the analysis of the events occurring in the ER [28, 31]. In the case of type III procollagen, the nucleation was found to be directed by the prolyl-4-hydroxylation of at least two Pro residues at the C-terminal end of the Gly-X-Pro repeats. Such Gly-X-Pro triplet repeats are also found at the C-terminal ends of the triple helical domains in other fibril-forming procollagens. Once the thermal stability of these sequences is increased by the prolyl 4-hydroxylation, the C-terminal triplet repeats start to wind around one another. This nucleation would be supported by an additional stabilizing effect of the adjacent trimerized C-propeptide domain, including the coiled-coil region. Simultaneous multi-site nucleation may be also inhibited by the interaction of ER-resident molecular chaperones.

3.2 Propagation of the Triple Helix

Once nucleated at the C-terminus of the Gly-X-Y repeats, the triple helices propagate toward the N-terminus. This process is often compared to fastening a zipper. However, it is still unclear whether this process actually does proceed as smoothly as fastening a zipper. The fastening of the triple helix is a relatively slow process compared to the folding of globular proteins. In the triple helical region comprised of Gly-X-Y repeats, the imino acids Pro and 4-hydroxyproline (Hyp) are most frequently found in the X and Y-positions, respectively. As peptide bonds connected with a Pro or a Hyp residue often take a *cis* configuration, the *cis*-peptide bonds of the proα-chains must isomerize into a *trans* configuration prior to or upon triple helix formation. This isomerization has been found to be the rate-limiting step in the propagation of the triple helix-forming process both in vitro and in vivo [13, 32, 33]. However, this step has been shown to be accelerated by the action of the peptidyl prolyl *cis-trans* isomerases (PPIases) that also functions as a molecular chaperone. The rate of triple helix-propagation may also be greatly affected by the local amino acid sequences [34–36]. It has been suggested that the propagation of the triple helix may occur in a punctuated fashion that involves local micro-unfolding.

In general, proteins possess sufficient but still marginal structural stability against thermal denaturation. This has long been suggested to be true for the thermal stability of collagen as well. For example, the melting temperature (Tm) of human type I collagen was determined to be 41–42 °C, which is only slightly above the human body temperature [34, 37, 38]. However, Leikina et al. have recently overturned this currently accepted paradigm by concluding that the triple helical region of type I collagen is actually thermally unstable at body temperature [39]. The authors analyzed the thermal unfolding of the triple helix by ultra-slow scanning calorimetry and isothermal circular dichroism and found that the thermal stability of triple helical human type I collagen is

below 36 °C, and that at 37 °C the collagen prefers to exist as random coils. This conclusion was supported by the recent observation that the equilibrium Tm value of pN type III collagen (collagen with N-propeptide) is also below the body temperature [40]. These findings indicate that there is a critical problem in the procollagen folding in homoisotherm cells, namely, the Gly-X-Y repeats of proα-chains do not possess an intrinsic ability to fold into triple helical conformation at body temperature, and the propagating triple helix is always confronted with the thermal destruction of the triple helix.

How, then, can the cell complete the thermodynamically challenged folding of procollagen? Bruckner et al. previously reported that the stability of triple helix is higher within the cell than in vitro and suggested that there may be a system in the ER that stabilizes the triple helix [41]. More recent evidence now suggests that ER-resident molecular chaperones, most likely Hsp47, may contribute to the stabilization of the triple helical folding intermediates of procollagen [42, 43]. This issue will also be touched on in the next section.

4 Molecular Chaperones Involved in Procollagen Biosynthesis

The folding, assembly and transport of proteins are often assisted by successive actions of molecular chaperones within the cell. The ER also possesses a set of molecular chaperones that are utilized in the biosynthesis of secretory proteins, lumenal ER proteins, and transmembrane proteins. The biosynthesis of pro-

Table 1 Molecular chaperones involved in collagen biosynthesis

Chaperone	Major binding-region	Comment
Bip/Grp78	C-Propeptide	Hsp70 homolog in the ER
Calnexin	C-Propeptide (N-linked sugar)	ER-resident lectin with a transmembrane domain
Calreticulin	C-Propeptide (N-linked sugar)	A calnexin homolog without the transmembrane domain
PDI	C-Propeptide	Also function as the β-subunit of P4-H
P4-H	Gly-X-Y repeats (single-chain)	Known as the principal modification enzyme for procollagen
PPIase	Gly-X-Y repeats (single-chain)	Some members of this chaperone may be involved in the procollagen folding
Hsp47	Gly-X-Y repeats (triple helical)	Collagen-specific heat-shock protein found in the ER

collagen molecules also requires the assistance of molecular chaperones residing in the ER. Supporting this is that procollagen is co-immunoprecipitated with most of the ER-resident chaperones [44, 45]. The ER chaperones involved in procollagen biosynthesis fall into two classes. One class includes the general ER-chaperones such as Bip/Grp78, Grp94, calnexin, calreticulin, protein disulfide isomerase (PDI), and the PPIases. These chaperones function with a variety of client proteins as they have broad binding specificities. Apart from the PPIases, this class of chaperones mainly binds to the non-collagenous (propeptide) domains of procollagen. The other class of chaperones includes the procollagen-specific molecular chaperones such as Hsp47 and P4-H, which has also been recognized to be a major procollagen-modifying enzyme. These specific chaperones interact with the collagenous domain comprised of the Gly-X-Y repeats and recent studies have revealed that these ER-resident chaperones are involved in the folding, cellular trafficking and quality control of procollagen molecules (Table 1, Fig. 4).

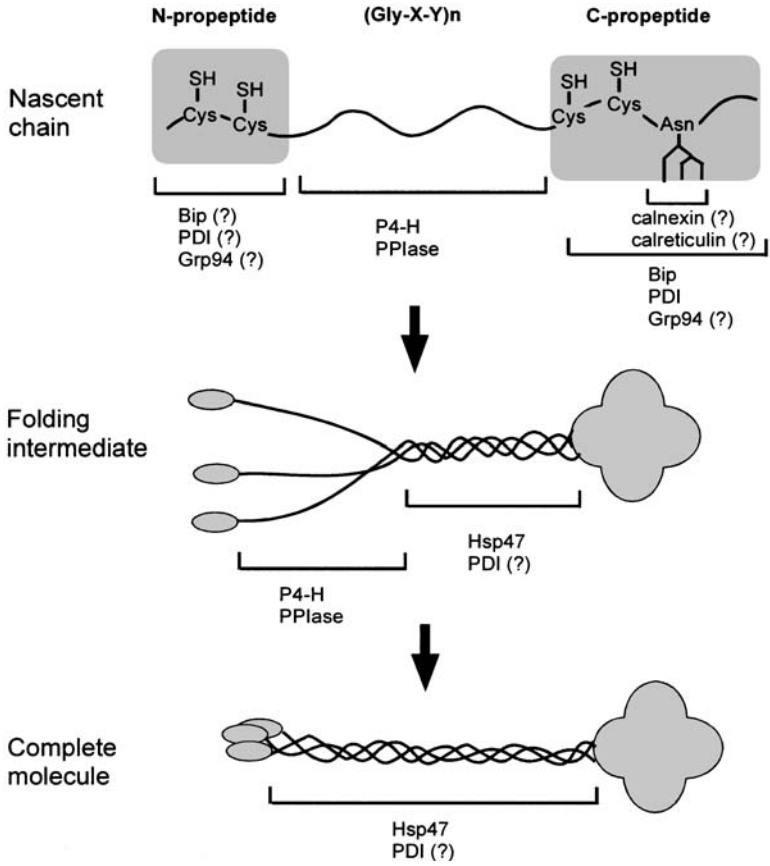

Fig. 4 Chaperone binding-sites on a procollagen molecule

The procollagen molecules have been shown to exist in an aggregated state and to form a reticular-like matrix in the ER lumen in vivo. The size of the aggregates can exceed 1500 kDa and they contain PDI [46]. This status of procollagen is also supported by the observations that the folding of the procollagen molecules occurs when the polypeptide chains are still in close association with the ER membrane, and that the membrane binding might be mediated by the ER chaperones [47]. These findings suggest that the folding of procollagen molecules may be accomplished with the assistance of a variety of ER-resident chaperones that form large chaperone complexes with the procollagen. In the following subsections, we will summarize the structures and functions of the individual ER-resident molecular chaperones that are involved in procollagen biosynthesis.

4.1 Bip/Grp78

Bip/Grp78 is the ER homologue of Hsp70 and bears the ER-retrieval signal Lys-Asp-Glu-Leu (KDEL) sequence at its C-terminal end. Bip, like other members of the Hsp70 family, shows broad binding specificity to short stretches of hydrophobic sequences [48]. Bip has been shown to interact stably with the mutated C-propeptide domain in osteogenesis imperfecta patients. In addition, the expression level of Bip in these patients was increased [18, 49]. This elevated expression of Bip is probably due to the activation of the unfolded protein response (UPR) pathway in the ER [50]. Consequently, it has been suggested that Bip is involved in the quality control mechanism that ensures only correctly-folded procollagen is transported to the Golgi apparatus, since it may capture procollagen molecules that have inappropriately-folded C-propeptide domains. However, it remains unclear whether Bip participates in the actual folding and quality control of procollagen in the ER.

4.2 Calnexin and Calreticulin

The C-propeptides of the α-chains of fibril-forming procollagens have the consensus sequence Asn-X-Thr/Ser for the addition of N-linked glycans. After the nascent proα-chain enters the ER lumen, N-linked oligosaccharide is attached to its C-propeptide region. This allows the proα-chain to be recognized by calnexin and/or calreticulin. Calnexin and calreticulin are homologous ER-resident lectins that recognize the monoglucose form of N-linked oligosaccharide of newly synthesized glycoproteins as molecular chaperones. The former is a membrane-bound protein whose lectin domain is oriented toward the ER lumen, and the latter is an ER-lumenal protein [51]. Although these lectin-type chaperones have been found to play critical roles in the folding and quality control of various glycoproteins, they have not been shown to function significantly in procollagen biosynthesis. Indeed, when the binding of the procolla-

gen C-propeptides to calnexin/calreticulin was abolished by mutation of the N-linked glycosylation sites, the assembly, folding, or secretion of the procollagen molecules did not appear to be affected [49].

4.3 Protein Disulfide Isomerase (PDI)

PDI is one of the most abundant proteins residing in the lumen of the ER. Mammalian PDI functions as a dimer of the 57 kDa subunit. PDI plays a central role in the folding of various secretory and membrane proteins that contain disulfide bridges by catalyzing the correct pairing of their disulfide bonds [52]. The disulfide exchange reaction is catalyzed by the formation of mixed disulfide bonds between the client proteins and PDI in redox-dependent manner [53]. PDI also transiently interacts with various polypeptides as a molecular chaperone to facilitate correct protein folding, even if they have no cysteine residues [54]. This chaperone activity of PDI has recently been reported to be independent from its redox state [55]. In addition, PDI exists as a component of multi-subunit proteins in the ER as it serves as the β-subunit of P4-H and as a small subunit in the microsomal triglyceride transfer protein.

The PDI polypeptide has a unique organization that consists of five structural domains, namely, a-b-b′-a′-c. The N-terminal four domains (a-b-b′-a′) have similar structural topology to that of thioredoxin [56, 57]. Of these, the highly homologous a and a′ domains contain a thioredoxin motif consisting of Cys-X-X-Cys that is responsible for catalyzing the thiol-disulfide exchange reaction. The non-catalytic b′ domain contains the principal client-binding site [58, 59]. The typical ER-retrieval signal Lys-Asp-Glu-Leu (KDEL) is found at the C-terminus of the c domain.

During procollagen biosynthesis, PDI interacts with procollagen either as the molecular chaperone or as a subunit of P4-H. PDI appears to associate with procollagen at its C-propeptide region [60, 61]. Since the trimerized forms of the C-propeptide domains of procollagens contain inter- and intramolecular disulfide bridges, the C-propeptide domains are likely to be substrates in PDI-catalyzed folding. In fact, it has been demonstrated that the binding of PDI to the C-propeptide plays a crucial role in coordinating the trimeric assembly of the proα-chains [60]. The individually expressed procollagen C-propeptide domains also remain associated with PDI and are retained in the ER until they are folded into native trimeric structures [61]. The stable interaction of PDI with C-propeptides that are unable to form trimers may also be involved in the quality control of procollagen, as has been described for Bip/Grp78.

It has also been found by utilizing the semi-permeabilized cell system that there is an additional interaction of PDI with fully folded type X collagen [62]. This interaction was not thiol-dependent and it was abolished when pH was lowered to pH 6.0, which is similar to the interaction between Hsp47 and collagen [63]. Unlike fibril-forming procollagen, the non-collagenous domains

of type X collagen are not enzymatically cleaved [64]. It may be that PDI helps to prevent the aggregation of type X collagen through its binding.

4.4 Prolyl 4-hydroxylase (P4-H)

P4-H is an essential enzyme in the post-translational modification of procollagen. The mammalian P4-Hs are heterotetramers that consist of two α subunits and two β subunits. The α subunit has an active site for the catalysis of prolyl 4-hydroxylation and three different subtypes of the α subunit have been identified [65–67]. The β subunit dimer of P4-H is identical to PDI. The β subunits are essential in the formation of an enzymatically active α2β2 tetramer, although they do not participate directly in the catalytic activity of prolyl hydroxylation. The catalytic function of the β subunit as PDI is not involved in the tetramerization of P4-H because the β subunits whose catalytic sites for PDI activity have been mutated can still form a α2β2 tetramer with full P4-H activity [68]. Thus, the P4-H β subunits may instead play a role in preventing the highly insoluble α subunit from forming inactive aggregates [69]. P4-H is localized in the lumen of the ER by virtue of the ER-retention signal in the β subunits. The expression level of the α subunit is lower than that of the β subunit, which allows the β subunits to act as PDI, independent of their function as P4-H.

The central role of P4-H in procollagen biosynthesis is its hydroxylation of the Pro residues at the Y positions of the Gly-X-Y repeats. This modification increases the thermal stability of the triple helix. Accumulating evidences suggest that P4-H also acts as a procollagen-specific molecular chaperone through its binding to Gly-X-Y repeats in single-stranded proα-chains. Recently, the substrate-binding domain in the P4-H α subunits was identified by partial enzymatic digestion of the tetrameric enzyme [70]. The substrate-binding domain was distinct from the catalytic domain. The single-chain collagenous sequences probably bind to the P4-H tetramer by interacting with this substrate-binding domain. Recognition of the proα-chains by this domain would facilitate the effective prolyl 4-hydroxylation of the long collagenous sequences. Although the hydroxylated model substrate $(Gly\text{-}Pro\text{-}Hyp)_5$ is a less effective binder to the domain than the non-hydroxylated peptide $(Pro\text{-}Pro\text{-}Gly)_5$ [71], the association of P4-H with single chain proα-chains has been observed in the semi-permeabilized cell system even after prolyl 4-hydroxylation has occurred [60]. As described above, the post-translational modifications of procollagen, including prolyl 4-hydroxylation, start before the translocation of the entire polypeptide is completed, and in the folding pathway, the triple helix-forming region of procollagen stays in an unfolded state until the proα-chains trimerize at the C-propeptide regions. The interactions of P4-H and even other modification enzymes with the triple helix-forming region of the single chain procollagens has been suggested to keep the chains in a folding-competent state, thus avoiding unfavorable nucleation and subsequent mis-aligned triple helix formation.

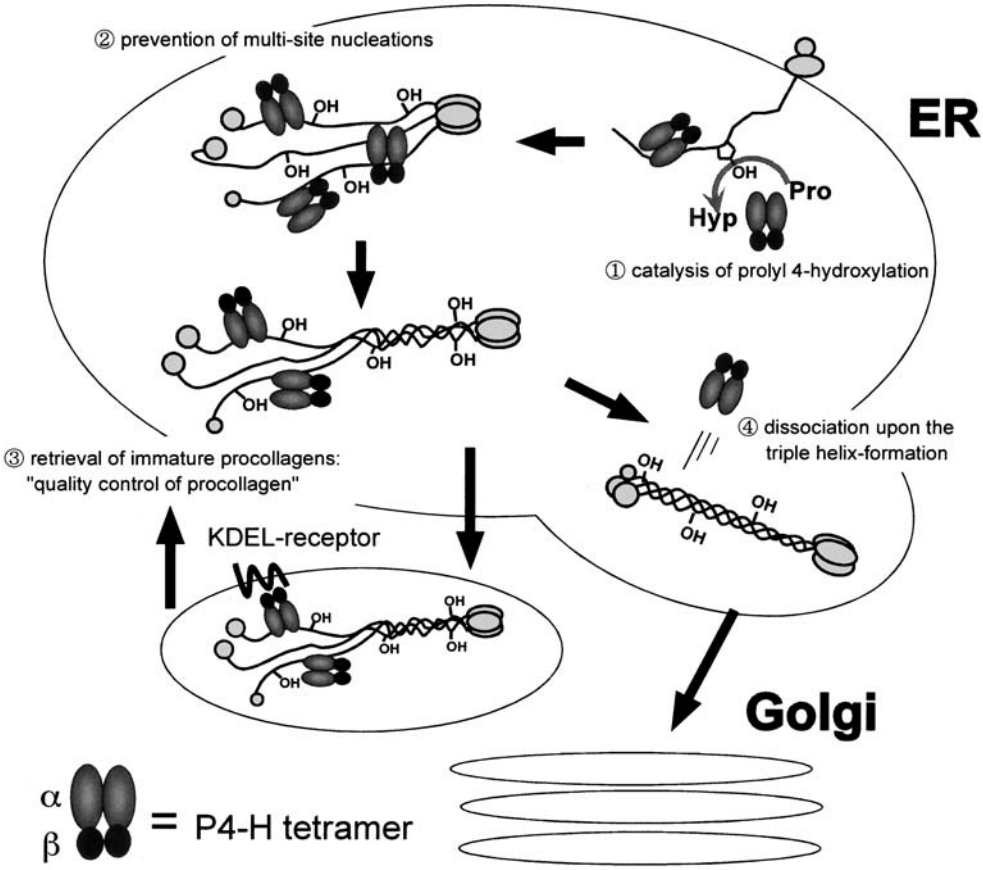

Fig. 5 Possible role of P4-H as a procollagen-specific molecular chaperone

In addition, the binding of P4-H to procollagen may contribute to the quality control of the procollagen molecules before they are transported to the Golgi apparatus since P4-H appears to capture the unfolded or incompletely folded procollagens that still have a single chain region, thus forcing their retention in the ER lumen until correct folding is accomplished [72]. Abnormal procollagen bearing genetic mutations or the deletion of the collagenous domain have also been shown to interact stably with P4-H and to be retained in the ER [73] (Fig. 5).

4.5 Peptidyl-Prolyl *cis-trans* Isomerase (PPIase)

The triple helix-forming region of procollagen contains large amounts of imino acid residues. In the Gly-X-Y repeats, about one-third of the X and Y positions are occupied by Pro and Hyp, respectively. Although peptide bonds in the *cis*

configuration are energetically unstable in most cases, the peptide bonds followed by a Pro or Hyp residue contain a considerable number of *cis*-conformers. It has been estimated that 15% of the Z-Pro and 8% of the Z-Hyp peptide bonds (Z denotes any amino acid residues) in nascent proα-chains of type I collagen are in the *cis*-configuration [74]. Prior to triple helix formation, these *cis*-peptide bonds in the proα-chains must be converted to *trans*-configurations. This *cis-trans* isomerization step is known to be the rate-limiting step in the propagation of the triple helix both in vitro [13, 33] and in vivo [75].

PPIases catalyze the *cis-trans* interconversion of prolyl peptide bonds in various proteins and are reported to have a chaperone-like activity [76, 77]. Members of the PPIase family are ubiquitously distributed in various organelles of various organisms. Proteins belonging to this family are classified into three subfamilies based on the different efficacies of small molecule inhibitors. Cyclophilin is a PPIase whose catalytic action is inhibited by the immunosupressant cyclosporin A. The activity of the members of the FKBP subfamily is inhibited by FK506, another immunosuppressant. The Parvulin family proteins are also included in the PPIase family. The ER of mammalian cells contains at least four different PPIases, namely, cyclophilin B, cyclophilin C, FKBP13 and FKBP65.

The importance of the PPIases in procollagen biosynthesis has been demonstrated by showing that inhibition of PPIase activity by cyclosporin A and FK506 delays the triple helix-formation of procollagen in cellulo [34, 78]. Although it is still unclear which ER-resident PPIase plays a central role in the triple helix formation of procollagen, at least two, cyclophilin B and FKBP65, have been shown to be major PPIases that directly interact with procollagen [79, 80].

4.6 Hsp47

Of the ER-resident molecular chaperones that are involved in procollagen biosynthesis, Hsp47 is one of the most important. Hsp47 was first identified as a collagen-binding heat-shock protein (HSP) with a molecular size of 47 kDa [81]. It is identical to the proteins that have been reported as colligin [82] and gp46 [83]. Hsp47 is a member of the serine protease inhibitor (SERPIN) superfamily but it does not show protease inhibitory activity. Hsp47 contains the ER retrieval signal Arg-Asp-Glu-Leu (RDEL) sequence at its C-termini and thus resides in the ER lumen [63, 84]. By using in vivo chemical crosslinking and immunoprecipitation techniques, Hsp47 has been revealed to associate with procollagen in the ER [44]. The interaction is transient and the procollagen molecules dissociate from Hsp47 upon or just before reaching the *cis*-Golgi [84].

Hsp47 was assumed to be a molecular chaperone that is specific for procollagen biosynthesis because of its transient association with procollagen in the ER in addition to the fact that it is a heat-inducible protein [85]. However, the real function of Hsp47 in procollagen biosynthesis has not been elucidated until recently. In the remainder of this section, we will summarize these recent

advances in our understanding of the role Hsp47 plays in procollagen biosynthesis.

In 2000, Nagai et al. succeeded in disrupting the *hsp47* gene in mice and demonstrated the indispensable role of Hsp47 in procollagen biosynthesis for the first time [86]. Although heterozygotic mice (*hsp47*+/–) showed no evident phenotype, homozygotic mice (*hsp47*–/–) showed an embryonic lethal phenotype; these mice died before 11.5 days post coitus (dpc) with severe impairment of collagen-based tissue structures, including reticular fibers in the mesenchyme and basement membranes (Fig. 6). Fibroblasts established from the *hsp47*–/– mice produced and secreted abnormal (probably incorrectly aligned) type I procollagen molecules into the medium. Embryonic stem (ES) cells where both alleles of the *hsp47* gene were disrupted also secreted insufficiently folded type IV collagens into the medium and did not produce basement membrane-like structures in the embryoid bodies [87]. It should be noted that the phenotype of the *hsp47*–/– mice is more severe than that of any other reported knockout mouse in which individual collagen proα-chains have been disrupted [88, 89]. These findings clearly suggest that the function of Hsp47 is not limited to a distinct type of collagen; rather, it may be that Hsp47 acts as a chaperone for various types of procollagens. This suggestion is supported by the observation that Hsp47 can interact with various types of collagens in vitro [90].

What is the function of Hsp47, and which step in collagen biosynthesis is facilitated by the action of Hsp47? Some clues to answer those questions were obtained from recent studies on the mechanism by which Hsp47 recognizes procollagen. It was not clear until recently which procollagen conformation is actually recognized by Hsp47. Earlier in vitro and in vivo studies indicated that

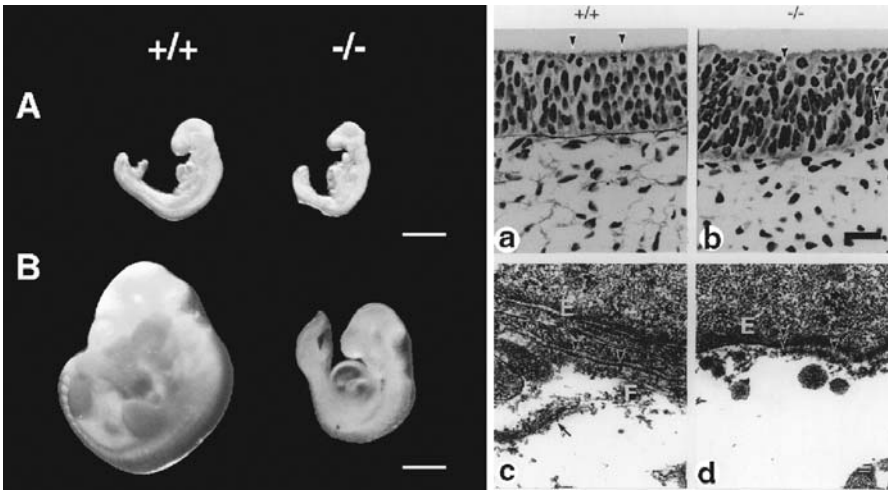

Fig. 6 Hsp47 null mice. **A** 9.5 dpc. **B** 10.5 dpc. **a**, **b** Neuroepithelium and underlying mesenchyme (9.5 dpc); **c**, **d** structures of basement membranes (9.5 dpc)

Hsp47 interacts with triple helical collagens [84, 90], while immunoprecipitation analyses indicated that the polysome-bound nascent proα-chain already binds to Hsp47 [45, 84]. Thus, Hsp47 appeared to interact with both single collagen chains and triple helical procollagen molecules. The conformation of procollagen that is recognized by Hsp47 has been recently appraised by using a more sophisticated strategy. Bulleid's group and our own have independently demonstrated by different strategies that Hsp47 preferentially recognizes the correctly folded triple helical conformation. The former group took advantage of the semi-permeabilized cell system and engineered a mini-procollagen with a truncated triple helical region. They then showed that only triple helical molecules were co-immunoprecipitated with Hsp47 [42]. In our studies, we selected the Hsp47-binding peptides in a random collagen-like peptide library that encodes $(\text{Gly-Pro-Y})_n$-repeats by using a yeast two-hybrid system with Hsp47 as a bait. Of the peptides that were selected by Hsp47, those whose Y-positions contained amino acids that stabilized the helix were positively selected [43]. In contrast, the peptides whose Y-positions were occupied by amino acids that destabilized the helix were not selected. Thus, the correctly folded triple helical conformation of procollagen is preferentially recognized by Hsp47. This makes Hsp47 unique as, in general, molecular chaperones recognize polypeptides with an as-yet non-native tertiary structure and are released from them after they have folded into their native (or correctly folded) structures. This general phenomenon is also the case for the other ER-chaperones that assist the proper folding of procollagen. As described in this section, the capturing and retention of unfolded and misfolded procollagen molecules by Bip, PDI, and P4-H is the key mechanism of the quality control system that ensures the generation of properly folded collagens. Thus, the conformational preference of Hsp47 constitutes a very unique client-recognition mechanism.

The sequence dependency of the interaction of Hsp47 with procollagen triple helices has also been elucidated recently. Koide et al. found that triple helical collagen model peptides that contain an Arg residue in the Y-position are specifically recognized by Hsp47 in vitro [91]. The interaction of Hsp47 with Arg-containing collagenous peptides is much stronger than its recognition of the previously identified Hsp47-binding collagen mimetic $(\text{Pro-Pro-Gly})_n$ [92]. The importance of Arg residues in types I and III collagen was also demonstrated by using chemically modified collagen. This revealed that Hsp47-binding was abolished only when the Arg residues on native collagen were specifically modified through 2,3-butanedione treatment as chemical modifications of other residues, including Lys, Glu, Asp and His, only slightly affected the Hsp47-collagen interaction [91]. This sequence specific interaction between procollagen and Hsp47 has also been demonstrated by using genetically engineered model procollagen in semi-permeabilized cells [93]. Based on the simple assumption that Hsp47 recognizes X-Arg-Gly sequences and that any adjacent sequences do not affect the binding, it has been estimated that type III collagen, for instance, bears 41 possible Hsp47 binding sites. However, more recent work on the in vitro interaction between Hsp47 and collagen has

suggested that there are additional structural requirements apart from the critical Arg residues, since only limited CNBr fragments of collagen could interact with Hsp47 [94]. At this stage, we cannot precisely estimate the stoichiometry of the Hsp47-procollagen complex but we believe that a procollagen molecule most probably possesses multiple Hsp47 binding sites.

Although more has been learned in the last few years about the mechanism by which Hsp47 recognizes procollagen, its function still remains enigmatic as well as how this molecule associates with and dissociates from procollagen. The following possible functions of Hsp47 have been proposed to date [95–97]:

1. Inhibition of intracellular procollagen degradation [98, 99].
2. Quality control of procollagen [92].
3. Control of procollagen transport to Golgi apparatus [100].
4. Stabilization of triple helical folding intermediates of procollagen [42, 43].
5. Inhibition of procollagen aggregate formation in the ER [101].

Conceptually, recent findings argue for the latter two functions, as will be discussed below.

As described above, the triple helical region of a collagen molecule prefers random coil structures at the equilibrium condition at 37 °C [39]. This finding strongly suggests the existence of a factor that stabilizes the triple helical portions in the folding intermediates of procollagen in the cells. Since Hsp47 preferentially binds to correctly folded triple helical portions of procollagen, this chaperone may be the best candidate as the triple helix stabilizer. This scenario is consistent with the fact that Hsp47 is induced by heat shock. All the ER-resident molecular chaperones have been reported to be induced by ER stress through the UPR pathway but only Hsp47 is induced by heat shock. Under heat-stress conditions, procollagen triple helix-formation should be more difficult, and the collagen triple helix would be destabilized. In such heat-stress conditions, the cells would require higher levels of Hsp47 to stabilize the procollagen triple helix. However, the likelihood of this mechanism is still being debated. Bächinger and co-workers have reported that the thermal stability and refolding rates of the types I and III collagen in vitro do not increase in the presence of Hsp47 [40]. Thus, this putative role Hsp47 may play in procollagen biosynthesis has to be corroborated by additional molecular studies.

Yet another probable function of Hsp47 is that it inhibits the aggregation of procollagen in the ER. Accordingly, it has been shown that the addition of Hsp47 prevents the self-association of the triple helix form of type I collagen in vitro [101]. Although it is not clear if this mechanism also works in the ER, there is supportive evidence that procollagen molecules tend to laterally associate in the Golgi apparatus, which is where the procollagens dissociate from Hsp47, and that this results in large molecular aggregates [10] (Fig. 7).

Of the ER-resident chaperones involved in procollagen biosynthesis, only Hsp47 binds to correctly folded procollagen. Moreover, even after correct folding has been completed, Hsp47 still remains bound to the procollagen in the

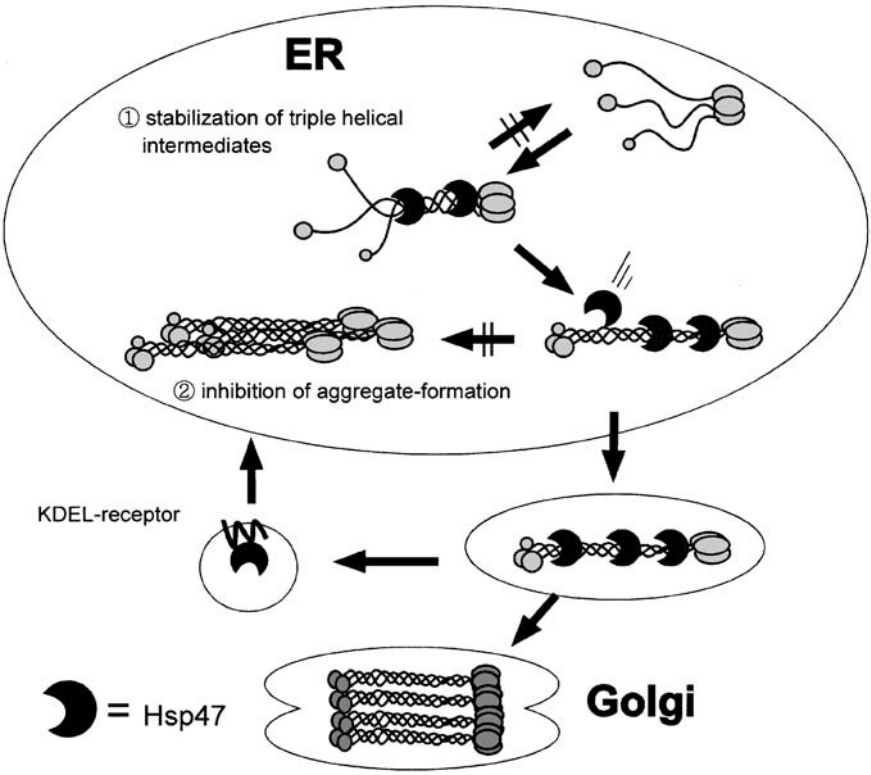

Fig. 7 Possible role of Hsp47

ER. How does Hsp47 dissociate from procollagen? It has been shown previously that Hsp47 dissociates from procollagen after leaving the ER, probably in the ER-Golgi intermediate compartment (ERGIC) or in the *cis*-Golgi [84]. A previous in vitro study also showed that the Hsp47-collagen interaction is sensitive to a change in the pH as Hsp47 dissociates from collagen in vitro when the pH is below 6.3 [63]. The lumenal pH of the pre-Golgi structures gradually decreases in parallel with their translocation to the Golgi region [102], and the pH in the trans-Golgi area has been reported to be even lower than pH 6. Taken together, these observations suggest that the dissociation of Hsp47 from procollagen may be caused by a change in the pH during its trafficking from the ER toward the Golgi apparatus. Alternatively, the dissociation may occur by the simple dilution of the procollagen-Hsp47 complex upon reaching the Golgi cisternae which harbor only a few free Hsp47 molecules [90].

In vertebrate tissues, the constitutive expression of Hsp47 shows a good positive correlation with that of the major types of collagens [85, 95, 103]. Cells producing higher levels of collagens also produce higher levels of Hsp47 and vice versa. In particular, the expression of Hsp47 is dramatically up-regulated under some pathophysiological conditions associated with collagen overpro-

duction or with abnormal accumulation of collagen. This correlation between collagen and Hsp47 expression has been suggested to be controlled at the transcriptional level, and *cis*-acting elements that are responsible for the tissue-specific expression of Hsp47 have been identified [104, 105]. These observations also support the notion that Hsp47 is an essential molecular chaperone for the proper biosynthesis of procollagen. However, this correlation between the expression of collagen and Hsp47 is not observed in invertebrates as Hsp47 has been found only in vertebrates higher than zebra fish. Genomic screens for Hsp47 protein or gene homologues in *Drosophila melanogaster* and *Caenohabititis elegans* have not been successful so far (Nagata et al., unpublished result), in spite of the fact that collagen is found in all multicellular animals. This suggests that Hsp47 may be a crucial component only in the vertebrate system of procollagen biosynthesis and that invertebrates follow a different scenario of procollagen production.

5
Intracellular Trafficking of Procollagen Molecules

5.1
ER-to-Golgi Transport

Secretory proteins folded and modified in the ER are generally secreted to the extracellular space via the Golgi apparatus. At the Golgi apparatus, proteins are further modified for the sorting to their final destinations. Once triple helix formation has been completed, procollagen molecules are also transported to the Golgi cisternae. Pre-Golgi trafficking of secretory proteins has been studied mainly using ts-045-G, a temperature-sensitive glycoprotein of the vesicular stomatitis virus. The vesicular coat complexes COPII and COPI play important roles in the protein trafficking between the ER and the Golgi apparatus [106–108]. Cargo proteins to be transported are first concentrated into buds coated by COPII, followed by the formation of nascent transport vesicles. These are subsequently coated with the COPI complex to form larger COPI-coated vesicles [109, 110]. The complexes then segregate into cargo-rich and COPI-rich domains. Only the latter returns to the ER, while the former is transported to the Golgi together with the cargo proteins.

The size of the classical transport vesicle that is generated by budding of ER-membrane is 60–80 nm in diameter, which may be too small to carry a procollagen molecule, which consists of a rigid 300 nm-long triple helical rod-like structure. Although the ER-to-Golgi transport of procollagen molecules has not been extensively investigated, Stephens and Pepperkok has recently shown that a distinct pathway is used to transport procollagen [111]. They analyzed the transport of procollagen in living cells by utilizing a genetically engineered procollagen in which green fluorescent protein (GFP) is fused to the C-terminal propeptide of proα1(I) chain. The transport complexes that contained procol-

lagen molecules were distinctively different from those containing other cargos, such as the ts-045-G protein, although procollagen exits the ER via COPII-coated exit sites. The procollagen transport complexes did not contain the ERGIC-53/p58 protein that all ER-to-Golgi transport complexes identified to date seem to carry. It has also been suggested that active processes are involved that concentrate the procollagen molecules into the transport vesicles upon ER budding. Electron microscopic observations have revealed the morphology of the procollagen-containing ER-to-Golgi transport complexes [10, 112], as tubular sack-like structures with a length exceeding 300 nm.

5.2 Intra- and Post-Golgi Transport

Once procollagen molecules reach the cisternae of the Golgi apparatus, the procollagen molecules self-associate to form aggregates approximately 320 nm in length and 170 nm in thickness [9, 10, 112]. These aggregates may result from the lateral packing of the rod-like procollagen molecules. Although how this aggregate formation occurs is not clear, the aggregate structure is reminiscent of the segment-long-spacing (SLS) aggregates of collagen with zero-D-arrayed molecular packaging that can be formed in vitro by the addition of ATP. In fact, it has been found that the procollagen and partially processed procollagen secreted by chick embryo tendon cells are similar in structure to the SLS aggregates [113]. It should be noted again that the *cis*-Golgi is an organelle where procollagen molecules are liberated from Hsp47 for the first time. The dissociation of Hsp47 may result in the aggregation of procollagen in the cis-Golgi. In general, the formation of large aggregates in cells is considered to be the consequence of dead-end folding and harmful to the cells. While the functional advantage of forming procollagen aggregates has not been determined as yet, one can speculate that it may be a way to protect procollagen molecules from attack from proteases in the Golgi apparatus such as matrix metalloproteases. It is also possible that procollagen, the individual triple helical domain of which does not possess sufficient thermal stability at body temperature [39], acquires thermal stability by forming the lateral aggregates.

The procollagen aggregates that form in the Golgi cisternae then move across the Golgi stacks without leaving the lumen of the Golgi cisternae [10]. It is still unclear whether this "cisternal maturation model" is applicable to the anterograde transport of small soluble proteins that is mediated by the COPII-vesicular transport system [114]. However, several observations, including electronmicroscopic analyses, suggest this model also applies to the anterograde transport between Golgi cisternae of general soluble cargos. After passing through the Golgi apparatus, the procollagen molecules are incorporated in secretory vesicles where the aggregates increase in length [115].

The procollagen molecules that have completed their folding and modifications are secreted into the extracellular space, probably in an aggregated form. During or following their secretion, the N- and C-propeptide domains are

cleaved from procollagens by specific processing enzymes, the N- and C-proteinases, and are incorporated into collagen fibrils (see Fig. 2).

6 Production of Recombinant Collagens

There is growing interest in collagen as a promising biomaterial in cosmetic surgery, tissue engineering and drug delivery systems [116, 117]. Most of the collagen currently used for these purposes is purified from the tissues of cows. The use of animal collagens is associated with the unavoidable risks of inducing gelatin allergies [118] or infection with prions that may cause bovine spongiform encephalopathy (BSE) [119]. Consequently, the production of human collagen, which is not associated with these risks, has been attempted in several laboratories. The information retrieved from these attempts has not only been useful for the future applications of recombinant human collagens but has also provided novel insights into the mechanism of procollagen biosynthesis.

To date, a number of different eukaryotic cells have been tested for the production of recombinant collagens, namely, budding [120, 121] and fission yeasts [122, 123], mammalian cell lines [124–126], and insect cells [127–130]. Transgenic animals and plants have also been candidate hosts for the production of recombinant human procollagen. It has been reported that recombinant procollagen can be purified from the milk of transgenic mice [131, 132], from cocoons of transgenic silkworms [133], and tobacco plants [134–136]. All of these hosts have been shown to produce triple helical collagen or procollagen molecules but their thermal stability may differ depending on the systems used. With the exception of the mammalian cell system, it was necessary to co-express P4-H subunits to produce properly hydroxylated and hence stable triple helical molecules. Moreover, although the production of collagen using the fission yeast *Pichia pastris* yielded as much as 1.1 g/l of collagen, the procollagen molecules were not secreted into the media [122]. This retention of procollagen within the cells has been also observed in other cell systems except for the mammalian cell lines. In addition, the systems using the mammary glands of transgenic mice and the silk glands of transgenic silkworms employ genetically engineered procollagen that has shortened triple helical domains. It has not been shown if these systems are capable of producing full-length procollagen. These observations emphasize the existence of a procollagen-specific intracellular trafficking system in innate procollagen-producing cells. In addition, as yeasts, insects, and plants are not likely to possess Hsp47, their inability in secreting procollagen molecules may relate to the lack of Hsp47. Consequently, it is tempting to speculate that Hsp47 may be involved in the procollagen-specific trafficking in innate procollagen-producing cells [137].

The *Pichia pastoris* expression system has also led to the discovery of an unusual control mechanism between P4-H and its procollagen substrate,

namely, that the formation of the stable α2β2 tetramer of P4-H requires the expression of procollagen and that the folding of procollagen requires the expression of the active tetrameric enzyme [122]. In addition, even more striking observations came from *Saccharomyces cerevisiae*, where human pro α-chains lacking the N- and C-propeptides are folded into correctly aligned triple helical molecules [121]. This unexpected observation challenges the consensus understanding that the C-propeptide domain is necessary for the initial association of the proα-chains (see above).

7 Conclusions and Future Perspectives

As described in this chapter, procollagen biosynthesis in cells constitutes a very complex system involving a number of unique processes that are generally not found in the synthesis and secretion of globular proteins. Recent progress has provided remarkable support to the widely accepted model of procollagen biosynthesis that is illustrated in Fig. 2. It should be noted that this progress has been made possible by applying novel techniques to investigate the events that occur during procollagen biosynthesis. One of these techniques is that of Bulleid and co-workers, who developed the system that utilizes the in vitro translation/translocation in semi-permeabilized cells [138]. This elegant system, along with the utilization of genetically engineered proα-chains, has made it possible to investigate the events that occur in the ER. This system has been used to extend our understanding of the functions of ER-resident molecular chaperones in procollagen folding. The development of a system for single-molecule imaging in living cells has also contributed to the study of the dynamic trafficking of procollagen between organelles [108]. This technology will enable us to obtain further insight into the intracellular trafficking of procollagen, including how procollagen quality is controlled and how the chaperones return to the ER.

The most striking finding in recent years is that the melting temperature of mammalian mature collagen, which is an indicator of the thermal stability of collagen molecules, is apparently below body temperature [39]. This finding has caused a paradigm shift in our understanding of the folding and stability of procollagen since the procollagen molecule cannot be expected to fold into its destined structure at the body temperature. Indeed, even after the structure is formed, the thermal unfolding may occur at least locally. This may be the reason why the procollagen molecule requires the Hsp47 chaperone, as it may specifically stabilize the triple helical assembly of procollagen. It may be also the reason why Hsp47 is the only heat-inducible stress protein in the ER of mammalian cells. Once procollagen molecules reach the *cis*-Golgi, they self-associate laterally and never dissociate thereafter [10].

Of all proteins that are integrally involved in the biosynthesis of procollagen, the function of Hsp47 seems to be the least clear. Indeed, the use of mammalian

and other model systems, including yeasts and invertebrates, to produce recombinant collagen have led to contradictory observations regarding the function of Hsp47. In mammals, Hsp47 is essential for proper procollagen biosynthesis while in some invertebrates, procollagen biosynthesis can be accomplished without Hsp47. It is thus possible to speculate that Hsp47 is a specialized chaperone that plays a role only in collagen-producing mammalian cells.

In this chapter, we focused on the procollagen synthesis in mammalian cells. However, multicellular animals ubiquitously synthesize collagen, and animals other than arthropods and nematodes have been shown (or predicted) to possess fibril-forming collagens [139]. Interestingly, Hsp47 has not been reported to be expressed in those animals. Do invertebrates possess similar machineries of procollagen synthesis or are their machineries completely different? We do not have enough information to answer the question, but we speculate that the system may vary from organism to organism. For instance, the collagen of the vent worm *Rifia pacyptila* that inhabits deep sea volcanic vents has been suggested to be stabilized by the addition of the di- and tri-saccharides of galactose to threonine residues in the Y-positions [140–142]. This type of collagen modification has been also suggested to occur in the earthworm *Lumbricus terrestris* and the clamworm *Nereis virens* [143, 144]. Thus, annelid collagens are likely to employ a different mechanism to stabilize the triple helices. Searching for the prototype of the mammalian system of procollagen production in lower organisms will provide further information for the mechanisms by which procollagen is synthesized and how these mechanisms have evolved. Moreover, it should be noted that bacteria, bacteriophages, and viruses possess genes encoding collagen-like molecules [145, 146]. Recent analyses suggested that triple helical collagen-like molecules containing tandem Gly-X-Y repeats can be synthesized in bacteria (Xu et al. 2002) which do not possess an ER suited to fold and modify eukaryotic procollagens. Analysis of the collagen biosynthetic pathways in such organisms would provide novel insights into the mammalian counterpart.

In the triple helix-formation of heterotrimeric procollagens, there is another question that has not yet been resolved. Within the triple helix, polypeptide chains are placed with one-residue-staggering along the helical axis. This would make it possible to form structural isomers of procollagen when it consists of different proα-chains. For instance, in the case of type I procollagen, the theoretical registers of the three chains along the helical axis could be $\alpha1\alpha1\alpha2$, $\alpha1\alpha2\alpha1$, and $\alpha2\alpha1\alpha1$. These hypothetical isomers of procollagen are expected to resemble each other in terms of their overall molecular structure and thermal stability. However, when looking more closely at these molecules, the surface profiles of their triple helices should differ. Does the cell selectively synthesize only one of the isomers? A clue to selective production has been obtained by using type IV collagen [147]. The residues (one Arg and two Asp residues) on type IV collagen, which are critical for $\alpha1\beta1$ integrin-binding, are located on different chains. By using the fluorescent energy transfer (FRET) technique followed by residue-specific fluorescent labeling, this region has been determined to have the register of $\alpha2$(IV) $\alpha1$(IV) $\alpha1$(IV). A subsequent study

was then conducted with a heterotrimeric collagen-model peptide with the corresponding integrin-binding sequence of type IV collagen [148]. As expected, the triple helical model with the native chain-register showed a higher affinity for α1β1 integrin, but the thermal stability of the natively-aligned triple helix was lower than the non-natively staggered one. This observation implies that the conformational stability of the natively staggered triple helix may not always, at least locally, be higher than that of mis-staggered ones.

As discussed in Section 6, the production of recombinant human collagens has been actively investigated in the hope of generating commercially available alternatives of animal collagens. However, most of the systems tested have failed, probably because of missing biosynthetic machineries in the respective hosts. The elucidation of the entire biosynthetic pathway that leads to mammalian collagen will greatly aid the production of recombinant collagens.

Control of collagen synthesis by effective drugs is a promising way to treat various fibrotic diseases, such as those of the liver, lung, kidney, and arteriosclerosis. All of the proteins that are specifically involved in procollagen biosynthesis, including the modifying enzymes and the chaperones, could be potential drug targets. To date, many compounds that inhibit P4-H activity have been identified and synthesized. Some of these, such as HOE 077 and Safironil, have been reported to effectively prevent the progression of fibrosis in vivo [149]. There are at least six P4-H enzymes in humans. Three are procollagen P4-Hs that show different tissue distributions and three are hypoxia-induced P4-Hs that control oxygen homeostasis [150]. As their mechanisms of enzyme action are very similar, the development of inhibitors that are specific for the P4-H isoenzymes is expected to be very useful in the treatment of fibrotic diseases. Inhibition of Hsp47 function may also be a promising objective of antifibrotic drugs, since expression of Hsp47 is elevated in various fibrotic diseases [151, 152]. Although small molecules that inhibit Hsp47 function have not yet been discovered, it has been shown that the administration of antisense oligonucleotides against Hsp47 suppresses the accumulation of collagen in experimental glomerulonephritis [153] and peritoneal fibrosis [154].

References

1. Hoppe HJ, Reid KB (1994) Protein Sci 3:1143
2. Hakansson K, Reid KB (2000) Protein Sci 9:1607
3. Krejci E, Coussen F, Duval N, Chatel JM, Legay C, Puype M, Vandekerckhove J, Cartaud J, Bon S, Massoulie J (1991) EMBO J 10:1285
4. Kodama T, Freeman M, Rohrer L, Zabrecky J, Matsudaira P, Krieger M (1990) Nature 343:531
5. Rohrer L, Freeman M, Kodama T, Penman M, Krieger M (1990) Nature 343:570
6. Prockop DJ, Kivirikko KI, Tuderman L, Guzman NA (1979) N Engl J Med 301:13
7. Kirk TZ, Evans JS, Veis A (1987) J Biol Chem 262:5540
8. Kivirikko K, Myllylä R, Pihlajaniemi T (1992) Post-translational modification of proteins. CRC Press

9. Trelstad RL, Hayashi K (1979) Dev Biol 71:228
10. Bonfanti L, Mironov AA Jr, Martinez-Menarguez JA, Martella O, Fusella A, Baldassarre M, Buccione R, Geuze HJ, Mironov AA, Luini A (1998) Cell 95:993
11. Schofield JD, Uitto J, Prockop DJ (1974) Biochemistry 13:1801
12. Olsen BR, Hoffmann H, Prockop DJ (1976) Arch Biochem Biophys 175:341
13. Bächinger HP, Bruckner P, Timpl R, Prockop DJ, Engel J (1980) Eur J Biochem 106:619
14. Doege KJ, Fessler JH (1986) J Biol Chem 261:8924
15. Niyibizi C, Eyre DR (1989) FEBS Lett 242:314
16. Kleman JP, Hartmann DJ, Ramirez F, van der Rest M (1992) Eur J Biochem 210:329
17. Mayne R, Brewton RG, Mayne PM, Baker JR (1993) J Biol Chem 268:9381
18. Chessler SD, Wallis GA, Byers PH (1993) J Biol Chem 268:18218
19. Pace JM, Kuslich CD, Willing MC, Byers PH (2001) J Med Genet 38:443
20. Dion AS, Myers JC (1987) J Mol Biol 193:127
21. Koivu J (1987) FEBS Lett 212:229
22. Zafarullah K, Brown EM, Kuivaniemi H, Tromp G, Sieron AL, Fertala A, Prockop DJ (1997) Matrix Biol 16:201
23. Bernocco S, Finet S, Ebel C, Eichenberger D, Mazzorana M, Farjanel J, Hulmes DJ (2001) J Biol Chem 276:48930
24. Hulmes DJ (2002) J Struct Biol 137:2
25. Lees JF, Tasab M, Bulleid NJ (1997) EMBO J 16:908
26. McAlinden A, Smith TA, Sandell LJ, Ficheux D, Parry DA, Hulmes DJ (2003) J Biol Chem 278:42200
27. Beck K, Brodsky B (1998) J Struct Biol 122:17
28. Bulleid NJ, Dalley JA, Lees JF (1997) EMBO J 16:6694
29. Kramer RZ, Bella J, Mayville P, Brodsky B, Berman HM (1999) Nat Struct Biol 6:454
30. Okuyama K, Arnott S, Takayanagi M, Kakudo M (1981) J Mol Biol 152:427
31. Bulleid NJ, Wilson R, Lees JF (1996) Biochem J 317:195
32. Buevich AV, Dai QH, Liu X, Brodsky B, Baum J (2000) Biochemistry 39:4299
33. Bächinger HP, Bruckner P, Timpl R, Engel J (1978) Eur J Biochem 90:605
34. Bächinger HP, Morris NP, Davis JM (1993) Am J Med Genet 45:152
35. Davis JM, Bächinger HP (1993) J Biol Chem 268:25965
36. Ackerman MS, Bhate M, Shenoy N, Beck K, Ramshaw JA, Brodsky B (1999) J Biol Chem 274:7668
37. Baker AT, Ramshaw JA, Chan D, Cole WG, Bateman JF (1989) Biochem J 261:253
38. Raghunath M, Bruckner P, Steinmann B (1994) J Mol Biol 236:940
39. Leikina E, Mertts MV, Kuznetsova N, Leikin S (2002) Proc Natl Acad Sci USA 99:1314
40. Macdonald JR, Bächinger HP (2001) J Biol Chem 276:25399
41. Bruckner P, Eikenberry EF (1984) Eur J Biochem 140:397
42. Tasab M, Batten MR, Bulleid NJ (2000) EMBO J 19:2204
43. Koide T, Aso A, Yorihuzi T, Nagata K (2000) J Biol Chem 275:27957
44. Nakai A, Satoh M, Hirayoshi K, Nagata K (1992) J Cell Biol 117:903
45. Ferreira LR, Norris K, Smith T, Hebert C, Sauk JJ (1994) J Cell Biochem 56:518
46. Kellokumpu S, Suokas M, Risteli L, Myllyla R (1997) J Biol Chem 272:2770
47. Beck K, Boswell BA, Ridgway CC, Bächinger HP (1996) J Biol Chem 271:21566
48. Blond-Elguindi S, Cwirla SE, Dower WJ, Lipshutz RJ, Sprang SR, Sambrook JF, Gething MJ (1993) Cell 75:717
49. Lamandé SR, Chessler SD, Golub SB, Byers PH, Chan D, Cole WG, Sillence DO, Bateman JF (1995) J Biol Chem 270:8642
50. Mori K (2000) Cell 101:451
51. Helenius A, Aebi M (2001) Science 291:2364

52. Freedman RB, Hirst TR, Tuite MF (1994) Trends Biochem Sci 19:331
53. Darby NJ, Creighton TE (1995) Biochemistry 34:11725
54. Cai H, Wang CC, Tsou CL (1994) J Biol Chem 269:24550
55. Lumb RA, Bulleid NJ (2002) EMBO J 21:6763
56. Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE (1996) Biochemistry 35:7684
57. Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE (1997) Curr Biol 7:239
58. Klappa P, Ruddock LW, Darby NJ, Freedman RB (1998) EMBO J 17:927
59. Pirneskoski A, Klappa P, Lobell M, Williamson RA, Byrne L, Alanen HI, Salo KE, Kivirikko KI, Freedman RB, Ruddock LW (2004) J Biol Chem 279:10374
60. Wilson R, Lees JF, Bulleid NJ (1998) J Biol Chem 273:9637
61. Bottomley MJ, Batten MR, Lumb RA, Bulleid NJ (2001) Curr Biol 11:1114
62. McLaughlin SH, Bulleid NJ (1998) Biochem J 331:793
63. Saga S, Nagata K, Chen WT, Yamada KM (1987) J Cell Biol 105:517
64. Kwan AP, Cummings CE, Chapman JA, Grant ME (1991) J Cell Biol 114:597
65. Helaakoski T, Vuori K, Myllyla R, Kivirikko KI, Pihlajaniemi T (1989) Proc Natl Acad Sci USA 86:4392
66. Annunen P, Helaakoski T, Myllyharju J, Veijola J, Pihlajaniemi T, Kivirikko KI (1997) J Biol Chem 272:17342
67. Van Den Diepstraten C, Papay K, Bolender Z, Brown A, Pickering JG (2003) Circulation 108:508
68. Vuori K, Pihlajaniemi T, Myllyla R, Kivirikko KI (1992) EMBO J 11:4213
69. John DC, Grant ME, Bulleid NJ (1993) EMBO J 12:1587
70. Myllyharju J, Kivirikko KI (1999) EMBO J 18:306
71. Hieta R, Kukkola L, Permi P, Pirila P, Kivirikko KI, Kilpelainen I, Myllyharju J (2003) J Biol Chem 278:34966
72. Walmsley AR, Batten MR, Lad U, Bulleid NJ (1999) J Biol Chem 274:14884
73. Chessler SD, Byers PH (1992) J Biol Chem 267:7751
74. Sarkar SK, Young PE, Sullivan CE, Torchia DA (1984) Proc Natl Acad Sci USA 81:4800
75. Bruckner P, Eikenberry EF (1984) Eur J Biochem 140:391
76. Kay JE (1996) Biochem J 314:361
77. Fisher G, Schmid F (1999) Molecular chaperones and folding catalysts. Academic Publishers, Amsterdam
78. Steinmann B, Bruckner P, Superti-Furga A (1991) J Biol Chem 266:1299
79. Smith T, Ferreira LR, Hebert C, Norris K, Sauk JJ (1995) J Biol Chem 270:18323
80. Zeng B, MacDonald JR, Bann JG, Beck K, Gambee JE, Boswell BA, Bächinger HP (1998) Biochem J 330:109
81. Nagata K, Saga S, Yamada KM (1986) J Cell Biol 103:223
82. Kurkinen M, Taylor A, Garrels JI, Hogan BL (1984) J Biol Chem 259:5915
83. Cates GA, Brickenden AM, Sanwal BD (1984) J Biol Chem 259:2646
84. Satoh M, Hirayoshi K, Yokota S, Hosokawa N, Nagata K (1996) J Cell Biol 133:469
85. Nagata K (1996) Trends Biochem Sci 21:22
86. Nagai N, Hosokawa M, Itohara S, Adachi E, Matsushita T, Hosokawa N, Nagata K (2000) J Cell Biol 150:1499
87. Matsuoka Y, Kubota H, Adachi E, Nagai N, Marutani T, Hosokawa N (2004) Mol Biol Cell 15:4467
88. Schnieke A, Harbers K, Jaenisch R (1983) Nature 304:315
89. Lohler J, Timpl R, Jaenisch R (1984) Cell 38:597
90. Natsume T, Koide T, Yokota S, Hirayoshi K, Nagata K (1994) J Biol Chem 269:31224
91. Koide T, Takahara Y, Asada S, Nagata K (2002) J Biol Chem 277:6178
92. Koide T, Asada S, Nagata K (1999) J Biol Chem 274:34523

93. Tasab M, Jenkinson L, Bulleid NJ (2002) J Biol Chem 277:35007
94. Thomson CA, Tenni R, Ananthanarayanan VS (2003) Protein Sci 12:1792
95. Nagata K (1998) Matrix Biol 16:379
96. Lamandé SR, Bateman JF (1999) Semin Cell Dev Biol 10:455
97. Hendershot LM, Bulleid NJ (2000) Curr Biol 10:R912
98. Jain N, Brickenden A, Ball EH, Sanwal BD (1994) Arch Biochem Biophys 314:23
99. Ball EH, Jain N, Sanwal BD (1997) Adv Exp Med Biol 425:239
100. Hosokawa N, Hohenadl C, Satoh M, Kühn K, Nagata K (1998) J Biochem (Tokyo) 124:654
101. Thomson CA, Ananthanarayanan VS (2000) Biochem J 349(3):877
102. Palokangas H, Ying M, Vaananen K, Saraste J (1998) Mol Biol Cell 9:3561
103. Nagata K (2003) Semin Cell Dev Biol 14:275
104. Hirata H, Yamamura I, Yasuda K, Kobayashi A, Tada N, Suzuki M, Hirayoshi K, Hosokawa N, Nagata K (1999) J Biol Chem 274:35703
105. Yasuda K, Hirayoshi K, Hirata H, Kubota H, Hosokawa N, Nagata K (2002) J Biol Chem 277:44613
106. Rothman JE, Wieland FT (1996) Science 272:227
107. Schekman R, Orci L (1996) Science 271:1526
108. Stephens DJ, Pepperkok R (2001) J Cell Sci 114:1053
109. Aridor M, Bannykh SI, Rowe T, Balch WE (1995) J Cell Biol 131:875
110. Shima DT, Scales SJ, Kreis TE, Pepperkok R (1999) Curr Biol 9:821
111. Stephens DJ, Pepperkok R (2002) J Cell Sci 115:1149
112. Leblond CP (1989) Anat Rec 224:123
113. Hulmes DJ, Bruns RR, Gross J (1983) Proc Natl Acad Sci USA 80:388
114. Glick BS, Malhotra V (1998) Cell 95:883
115. Trelstad RL (1971) J Cell Biol 48:689
116. Lee CH, Singla A, Lee Y (2001) Int J Pharm 221:1
117. Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M, Polarek J (2003) Adv Drug Deliv Rev 55:1547
118. Sakaguchi M, Hori H, Hattori S, Irie S, Imai A, Yanagida M, Miyazawa H, Toda M, Inouye S (1999) J Allergy Clin Immunol 104:695
119. O'Grady JE, Bordon DM (2003) Adv Drug Deliv Rev 55:1699
120. Toman PD, Chisholm G, McMullin H, Giere LM, Olsen DR, Kovach RJ, Leigh SD, Fong BE, Chang R, Daniels GA, Berg RA, Hitzeman RA (2000) J Biol Chem 275:23303
121. Olsen DR, Leigh SD, Chang R, McMullin H, Ong W, Tai E, Chisholm G, Birk DE, Berg RA, Hitzeman RA, Toman PD (2001) J Biol Chem 276:24038
122. Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI (1997) EMBO J 16:6702
123. Myllyharju J, Nokelainen M, Vuorela A, Kivirikko KI (2000) Biochem Soc Trans 28:353
124. Geddis AE, Prockop DJ (1993) Matrix 13:399
125. Fertala A, Sieron AL, Ganguly A, Li SW, Ala-Kokko L, Anumula KR, Prockop DJ (1994) Biochem J 298:31
126. Fichard A, Tillet E, Delacoux F, Garrone R, Ruggiero F (1997) J Biol Chem 272:30083
127. Lamberg A, Helaakoski T, Myllyharju J, Peltonen S, Notbohm H, Pihlajaniemi T, Kivirikko KI (1996) J Biol Chem 271:11988
128. Myllyharju J, Lamberg A, Notbohm H, Fietzek PP, Pihlajaniemi T, Kivirikko KI (1997) J Biol Chem 272:21824
129. Tomita M, Ohkura N, Ito M, Kato T, Royce PM, Kitajima T (1995) Biochem J 312:847
130. Tomita M, Kitajima T, Yoshizato K (1997) J Biochem (Tokyo) 121:1061
131. John DC, Watson R, Kind AJ, Scott AR, Kadler KE, Bulleid NJ (1999) Nat Biotechnol 17:385
132. Bulleid NJ, John DC, Kadler KE (2000) Biochem Soc Trans 28:350

133. Tomita M, Munetsuna H, Sato T, Adachi T, Hino R, Hayashi M, Shimizu K, Nakamura N, Tamura T, Yoshizato K (2003) Nat Biotechnol 21:52
134. Ruggiero F, Exposito JY, Bournat P, Gruber V, Perret S, Comte J, Olagnier B, Garrone R, Theisen M (2000) FEBS Lett 469:132
135. Perret S, Merle C, Bernocco S, Berland P, Garrone R, Hulmes DJ, Theisen M, Ruggiero F (2001) J Biol Chem 276:43693
136. Merle C, Perret S, Lacour T, Jonval V, Hudaverdian S, Garrone R, Ruggiero F, Theisen M (2002) FEBS Lett 515:114
137. Tomita M, Yoshizato K, Nagata K, Kitajima T (1999) J Biochem (Tokyo) 126:1118
138. Wilson RR, Bulleid NJ (2000) Methods Mol Biol 139:1
139. Boot-Handford RP, Tuckwell DS (2003) Bioessays 25:142
140. Bann JG, Bächinger HP (2000) J Biol Chem 275:24466
141. Bann JG, Peyton DH, Bächinger HP (2000) FEBS Lett 473:237
142. Bann JG, Bächinger HP, Peyton DH (2003) Biochemistry 42:4042
143. Muir L, Lee YC (1970) J Biol Chem 245:502
144. Spiro RG, Bhoyroo VD (1980) J Biol Chem 255:5347
145. Rasmussen M, Jacobsson M, Bjorck L (2003) J Biol Chem 278:32313
146. Xu Y, Keene DR, Bujnicki JM, Hook M, Lukomski S (2002) J Biol Chem 277:27312
147. Golbik R, Eble JA, Ries A, Kühn K (2000) J Mol Biol 297:501
148. Saccá B, Sinner EK, Kaiser J, Lubken C, Eble JA, Moroder L (2002) Chembiochem 3:904
149. Wang YJ, Wang SS, Bickel M, Guenzler V, Gerl M, Bissell DM (1998) Am J Pathol 152:279
150. Hirsila M, Koivunen P, Gunzler V, Kivirikko KI, Myllyharju J (2003) J Biol Chem 278: 30772
151. Masuda H, Fukumoto M, Hirayoshi K, Nagata K (1994) J Clin Invest 94:2481
152. Razzaque MS, Taguchi T (1999) Histol Histopathol 14:1199
153. Sunamoto M, Kuze K, Tsuji H, Ohishi N, Yagi K, Nagata K, Kita T, Doi T (1998) Lab Invest 78:967
154. Nishino T, Miyazaki M, Abe K, Furusu A, Mishima Y, Harada T, Ozono Y, Koji T, Kohno S (2003) Kidney Int 64:887

Top Curr Chem (2005) 247: 115–147
DOI 10.1007/b103821
© Springer-Verlag Berlin Heidelberg 2005

Intracellular Post-Translational Modifications of Collagens

Johanna Myllyharju (✉)

Collagen Research Unit, Biocenter Oulu and Department of Medical Biochemistry and Molecular Biology, University of Oulu, 90014 Oulu, Finland
johanna.myllyharju@oulu.fi

1 Introduction 116

2 Occurrence and Functions of 4-Hydroxyproline, 3-Hydroxyproline and Hydroxylysine in Animal Proteins 117

3 Collagen Hydroxylases 120
3.1 Collagen Prolyl 4-Hydroxylases 120
3.1.1 Vertebrate Collagen Prolyl 4-Hydroxylases 120
3.1.2 Non-Vertebrate Collagen Prolyl 4-Hydroxylases 125
3.2 Lysyl Hydroxylase 127
3.2.1 Vertebrate Lysyl Hydroxylases 127
3.2.2 *Caenorhabditis elegans* Lysyl Hydroxylase 129
3.3 Prolyl 3-Hydroxylases 130
3.4 Peptide Substrates of Collagen Hydroxylases 130
3.5 Cosubstrates and Reaction Mechanisms 132
3.6 Inhibitors and Inactivators 137

4 Collagen Glycosyltransferases 141

References 142

Abstract Collagen synthesis involves many post-translational modifications, the collagen-specific intracellular modifications consisting of hydroxylation of certain proline and lysine residues to 4-hydroxyproline, 3-hydroxyproline and hydroxylysine, and glycosylation of some of the hydroxylysine residues to galactosylhydroxylysine and glucosylgalactosylhydroxylysine. The five enzymes catalyzing these modifications are collagen prolyl 4-hydroxylase, prolyl 3-hydroxylase, lysyl hydroxylase, collagen galactosyltransferase and collagen glucosyltransferase, all residing within the lumen of the endoplasmic reticulum. Vertebrate collagen prolyl 4-hydroxylase, prolyl 3-hydroxylase and lysyl hydroxylase have at least three isoenzymes, and all collagen hydroxylases belong to the group of 2-oxoglutarate dioxygenases, which require Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate. Although the three-dimensional structures of these enzymes are still unknown, detailed information is available on their catalytically critical residues and reaction mechanisms. 4-Hydroxyproline residues have an essential role in providing the collagen triple helices with thermal stability, and collagen prolyl 4-hydroxylase is therefore regarded as an attractive target for chemical inhibition to control excessive collagen accumulation, e.g. in fibrotic diseases and cases of severe scarring.

Keywords Collagen · Collagen prolyl 4-hydroxylase · Prolyl 3-hydroxylase · Lysyl hydroxylase · Collagen glycosyltransferase

1 Introduction

Collagen synthesis involves an unusually large number of co-translational and post-translational modifications, many of which are unique to collagens and other proteins with collagen-like domains. The aim of this chapter is to review current information on the collagen-specific enzymes involved in the intracellular modification steps, which include hydroxylation of certain proline and lysine residues to 4-hydroxyproline, 3-hydroxyproline and hydroxylysine, and glycosylation of some of the hydroxylysine residues to galactosylhydroxylysine and glucosylgalactosylhydroxylysine (Fig. 1) [1–4]. These steps are catalyzed by

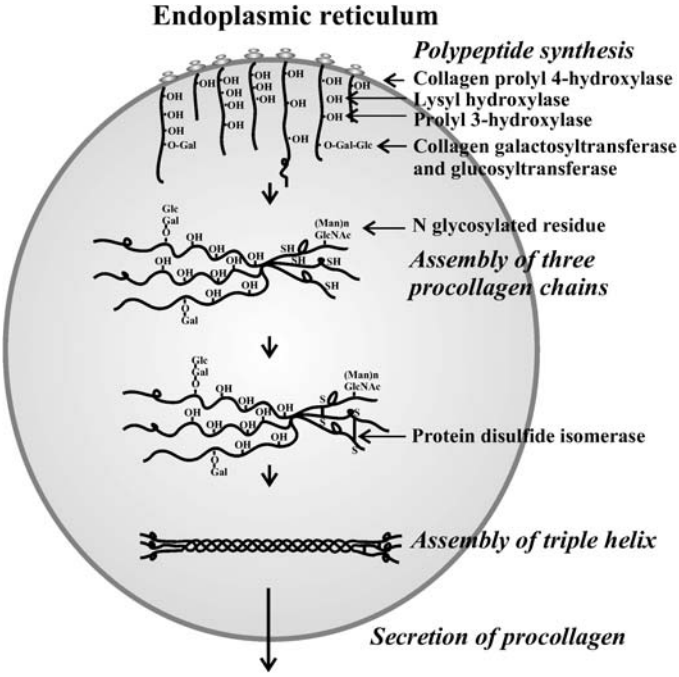

Fig. 1 The main intracellular steps in the synthesis of a fibril-forming collagen. The collagen chains are synthesized on membrane-bound ribosomes and secreted into the lumen of the endoplasmic reticulum. The collagen-specific modifications include hydroxylation of certain proline and lysine residues to 4-hydroxyproline, 3-hydroxyproline and hydroxylysine, and glycosylation of some of the hydroxylysine residues to galactosylhydroxylysine and glucosylgalactosylhydroxylysine. Certain asparagine residues in the C propeptides, or both the N and C propeptides, are also glycosylated by reactions similar to those in many other proteins. Association of the three collagen chains is directed in a type-specific manner by recognition sequences in their C propeptides, and the formation of intrachain and interchain disulfide bonds is catalyzed by protein disulfide isomerase. The triple-helical domain then nucleates at its C-terminal end and the triple helix is propagated in a zipper-like fashion towards the N terminus

five specific enzymes: collagen prolyl 4-hydroxylase, prolyl 3-hydroxylase, lysyl hydroxylase, collagen galactosyltransferase and collagen glucosyltransferase, all residing within the lumen of the endoplasmic reticulum (ER) [1–4]. The collagen hydroxylases, i.e. collagen prolyl 4-hydroxylase, prolyl 3-hydroxylase and lysyl hydroxylase, catalyze the formation of 4-hydroxyproline, 3-hydroxyproline and hydroxylysine almost exclusively in -X-Pro-Gly-, -Pro-4HyP-Gly- and -X-Lys-Gly-sequences, respectively (the repeating -Gly-X-Y-sequences that are typical of collagens and collagen domains in other proteins are written here as -X-Y-Gly- because of the hydroxylation properties of the three hydroxylases; see below) [5, 6]. Collagen prolyl 4-hydroxylase and lysyl hydroxylase were cloned about 15 years ago, while prolyl 3-hydroxylase was cloned only in 2004, all the human enzymes having at least three isoenzymes [4–8]. Collagen galactosyltransferase and collagen glucosyltransferase have not been cloned yet, although one of the lysyl hydroxylase isoenzymes has also been shown to possess low amounts of the collagen glycosyltransferase activities [9–11].

2 Occurrence and Functions of 4-Hydroxyproline, 3-Hydroxyproline and Hydroxylysine in Animal Proteins

Most of the 4-hydroxyproline, 3-hydroxyproline and hydroxylysine residues in animal proteins are found in the -X-4Hyp-Gly-, -3Hyp-4Hyp-Gly- and -X-Hyl-Gly- sequences of collagens. 4-Hydroxyproline, and hydroxylysine in most but not all cases, are also present in more than 20 additional proteins with collagen-like triple-helical domains, including the subcomponent C1q of complement, a C1q-like factor, adiponectin, at least eight collectins and three ficolins (humoral lectins of the innate immune defence system), the tail structure of acetylcholinesterase, three macrophage receptors, ectodysplasin, two EMILINS (elastic fibre-associated glycoproteins) and a src-homologous-and-collagen protein [3, 4, 6].

4-Hydroxyproline residues have an essential role in providing the collagen triple helices with thermal stability. The denaturation temperature of a non-hydroxylated type I collagen is only 24 °C, while a triple helix consisting of fully hydroxylated collagen polypeptide chains is stable up to 39 °C [12, 13]. Non-hydroxylated collagen polypeptide chains thus cannot form functional triple-helical molecules in vivo, and almost complete hydroxylation of the Y-position proline residues of the -X-Y-Gly- triplets is required for the generation of a collagen molecule that is stable at 37 °C. The stabilizing effect of 4-hydroxyproline residues on the collagen triple helix is most likely due to the inductive effects of the electron-withdrawing hydroxy group of 4-hydroxyproline residues on the pyrrolidine ring pucker, the θ and ψ torsional angles and the peptide bond *trans*/*cis* ratio of the substituted proline [14]. The hydroxylysine residues of collagens have two important functions: their hydroxyl groups serve as attach-

ment sites for carbohydrate units, and they are essential for the stability of the intermolecular cross-links [6]. The function of 3-hydroxyproline residues is not yet well understood, but it has recently been suggested that they may modulate the stability of the triple helix, allowing for specific regions of lower stability that may be necessary for the assembly of certain supramolecular structures, e.g. the meshwork structure formed by type IV collagen in basement membranes [15, 16].

Various collagen types show only relatively small, but distinct differences in their 4-hydroxyproline content [17]. A typical example is the most abundant collagen, type I, which contains about 100 4-hydroxyproline residues per 1000 amino acids, i.e. about 50% of the incorporated proline residues are hydroxylated [17]. In contrast, the 3-hydroxyproline and hydroxylysine contents of different collagen types vary markedly, the values ranging from 0 to over 10 residues per 1000 amino acids in the case of 3-hydroxyproline, and from about 5 to 70 in that of hydroxylysine [17]. Further variations in the amounts of 4-hydroxyproline, 3-hydroxyproline and hydroxylysine are found within the same collagen type in different tissues and even in the same tissue in various physiological and pathological states [6, 17]. Because of the critical function of the 4-hydroxyproline residues, the amount of this residue in a certain collagen type varies only within narrow limits, whereas large variations are found in 3-hydroxyproline and hydroxylysine. The 3-hydroxyproline content of type IV collagen, for example, can vary from about 1 to 20 per 1000 residues, and the hydroxylysine content of type I collagen from about 6 to 17/1000 [17].

4-Hydroxyproline is also found in -Gly-X-Y- sequences of elastin, the main component of elastic fibres [6, 17]. Elastin differs from collagens and proteins with collagen-like domains, however, in that its -Gly-X-Y- repeats do not form triple-helical structures. Although 4-hydroxyproline and hydroxylysine were long thought to be almost exclusively present in the Y positions of repeating -X-Y-Gly- sequences (the repeating collagenous -Gly-X-Y- sequences are again written here as -X-Y-Gly- because of the hydroxylation properties of the corresponding hydroxylases; see below), some exceptions were already recognized quite early on. 4-Hydroxyproline has been identified in single -X-4Hyp-Ala- triplets in the α3 chain of type IV collagen and in two of the polypeptide chains of the subcomponent C1q of human complement, and the telopeptides in some fibril-forming collagens contain hydroxylysine in one -X-Hyl-Ala- or -X-Hyl-Ser- sequence [6, 17]. In addition, 4-hydroxyproline and hydroxylysine are found in some proteins that contain a single -X-Y-Gly- sequence. For example, the kinin mixture in human plasma, urine and ascitic fluid contains both lysyl bradykinin and small amounts of hydroxyproline-lysyl-bradykinin with a single 4-hydroxyproline residue in the sequence -Pro-4Hyp-Gly- [18, 19]. A single 4-hydroxyproline residue is also found in an -Arg-4Hyp-Gly- sequence in the hydroxyproline luteinizing hormone-releasing hormone [20]. Correspondingly, anglerfish somatostatin-28 has a single hydroxylysine residue in the sequence -Trp-Hyl-Gly- [21, 22].

Most invertebrate collagens resemble the vertebrate collagens and have roughly similar 4-hydroxyproline contents in their triple-helical domains [6, 17]. Earthworm cuticle collagen is unique among all collagens studied so far, however, in that more than 90% of the incorporated proline residues are found in the form of 4-hydroxyproline, while the other extreme is represented by *Ascaris* cuticle collagen in which only about 5% of the proline residues are hydroxylated [6, 17]. Furthermore, most of the 4-hydroxyproline residues in earthworm cuticle collagen are found in the X positions of the repeating -X-Y-Gly- sequences [6, 17].

3-Hydroxyproline has been identified in collagens only in the sequence -Gly-3Hyp-4Hyp-Gly-. It has also been found in several proteins of the parasitic trematode *Fasciola hepatica*, including secreted cathepsin L-like proteinases, but not collagen. The 3-hydroxyproline residues of the *F. hepatica* proteins are present in sequences showing no consensus amino acid sequence [23, 24].

4-Hydroxyproline residues have very recently been identified in the hypoxia-inducible transcription factor HIF, where they have been shown to have a novel function in the means by which HIF controls gene expression in response to changes in the amount of cellular oxygen [25, 26]. HIF is an αβ heterodimer in which the stability of the α subunit is regulated in an oxygen-dependent manner. The HIF-α subunit is synthesized continuously, and one or both critical proline residues in two -Leu-X-X-Leu-Ala-Pro- sequences are hydroxylated under normoxic conditions [25, 26]. Hydroxylation of HIF-α is not catalyzed by collagen prolyl 4-hydroxylases [26] but by a novel cytoplasmic HIF prolyl 4-hydroxylase family [27–29], the resulting 4-hydroxyproline residue being essential for the binding of HIF-α to the von Hippel-Lindau E3 ubiquitin ligase complex and for rapid subsequent proteasomal degradation in normoxia [25, 26]. Under hypoxic conditions the oxygen-requiring hydroxylation step is prevented, HIF-α escapes degradation and dimerizes with HIF-β. The dimer is then translocated into the nucleus and becomes bound to the HIF-responsive elements in a number of hypoxia-inducible genes, such as those for erythropoietin, vascular endothelial growth factor and glycolytic enzymes [25, 26]. A striking difference in the catalytic properties of the HIF and collagen prolyl 4-hydroxylases is that the K_m values of the HIF prolyl 4-hydroxylases for oxygen are very high, even slightly above the concentration of dissolved O_2 in air, while the K_m values of collagen prolyl 4-hydroxylases are about one-sixth of these [30]. These data are consistent with the functions of the two classes of prolyl 4-hydroxylase. The HIF prolyl 4-hydroxylases are effective oxygen sensors that are inhibited by even a small decrease in O_2 concentration, whereas the collagen prolyl 4-hydroxylases must be able to function in situations with low O_2 concentrations, as in wounds and tissues of low vascularity.

3 Collagen Hydroxylases

3.1 Collagen Prolyl 4-Hydroxylases

3.1.1 Vertebrate Collagen Prolyl 4-Hydroxylases

Collagen prolyl 4-hydroxylase was first purified to near homogeneity more than 30 years ago from chick embryos by conventional procedures. The development of two affinity purification methods and various recombinant expression systems has subsequently facilitated the isolation of large quantities of pure enzyme [5–7, 31].

The collagen prolyl 4-hydroxylases from all vertebrate sources studied so far are $\alpha_2\beta_2$ tetramers (Fig. 2) in which the two catalytic sites are located in the α subunits, while the β subunits are identical to the multifunctional enzyme and chaperone protein disulfide isomerase (PDI) [5–7]. The molecular weight of the collagen prolyl 4-hydroxylase tetramer is about 240 kDa, those of the α and β subunits being about 63 kDa and 58 kDa, respectively [5–7]. The monomeric subunits possess no hydroxylase activity, and all attempts to assemble an active enzyme tetramer from the dissociated monomers in vitro have been unsuccessful [5–7]. Active recombinant collagen prolyl 4-hydroxylases have been successfully produced in insect [32], yeast [33] and plant cells [34], however, by coexpressing the α and β subunits. This has made it possible to study the molecular and functional properties of various collagen prolyl 4-hydroxylases in detail and to develop high-level recombinant expression systems for hydroxylated human collagens besides in mammalian cells also in insect cells, yeasts and plants [33–38].

In addition to its critical role in the hydroxylation of collagen chains, and thus in the generation of stable collagen triple helices, collagen prolyl 4-hydroxylase also plays an important role in collagen molecule quality control. The

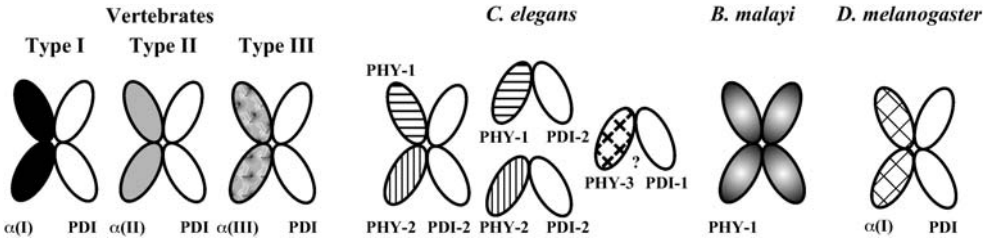

Fig. 2 Schematic representation of the forms of collagen prolyl 4-hydroxylase characterized in various species. The molecular composition of the active enzyme formed by the *C. elegans* PHY-3 and *C. elegans* PDI-1 is currently unknown

enzyme forms a stable association with non-helical collagen chains that is not dependent on the extent of their hydroxylation and retains the non-helical trimers within the ER [39]. Once the folding of the triple-helical domain is complete, the enzyme dissociates rapidly and releases the molecule for further transport [39].

3.1.1.1
The Catalytic α Subunit

The vertebrate α subunit was first cloned from human [40] and chicken [41] in 1989, and subsequently from the rat [42] and mouse [43]. Collagen prolyl 4-hydroxylase was long assumed to be of one type only, with no isoenzymes, but two additional α subunit isoforms, designated α(II) and α(III), have now been cloned and characterized from human and rodent sources [43–46]. Correspondingly, the α subunit first identified is now called α(I). The α(II) and α(III) subunits also become assembled into $\alpha_2\beta_2$ tetramers with PDI (Fig. 2), and the $[\alpha(I)]_2\beta_2$, $[\alpha(II)]_2\beta_2$ and $[\alpha(III)]_2\beta_2$ tetramers are referred to as type I, II and III collagen prolyl 4-hydroxylases, respectively [43–46]. Insect cell coexpression experiments have suggested that it is highly unlikely that vertebrate α subunits would form mixed tetramers containing two kinds of catalytic subunit [44].

The type I collagen prolyl 4-hydroxylase is the main form in most vertebrate cell types and tissues, but the type II enzyme is the major form in chondrocytes, osteoblasts, endothelial cells and cells in epithelial structures [47, 48]. The type II enzyme represents at least 70% and 80% of the total prolyl 4-hydroxylase activity in cultured mouse chondrocytes and cartilage, respectively [47], and it may thus have a major role in the development of cartilage and cartilagenous bone. The type III collagen prolyl 4-hydroxylase is expressed in many human tissues, but at much lower levels than the type I and type II enzymes [45].

The human α(I), α(II) and α(III) subunits consist of 517, 514 and 525 residues, respectively, with signal peptides of additional 17, 21 and 19 residues (Fig. 3) [40, 44–46]. The overall amino acid sequence identity between the processed human α(I) and α(II) subunits is 65%, and those between α(I) and α(III) and between α(II) and α(III) are 35–37% [44, 45]. The highest degree of identity is seen within the catalytically important C-terminal region [5–7, 49], the 120 C-terminal residues of the human α(I) subunit being 80% identical to those of human α(II), while α(III) is 56–57% identical to α(I) and α(II) in this region [44, 45]. All four critical residues at the catalytic site, which will be discussed in more detail below, are conserved in all three α subunits (Fig. 3).

The peptide-substrate-binding domain of the collagen prolyl 4-hydroxylases is distinct from the catalytic C-terminal domain and is located between residues Phe144 and Ser244 in the human α(I) subunit (Fig. 3) [50]. Recent nuclear magnetic resonance, surface plasmon resonance and isothermal titration calorimetry studies have indicated that many characteristic features of the binding of peptides to collagen prolyl 4-hydroxylases (see below) can be explained by the properties of binding to this domain rather than the catalytic domain [51].

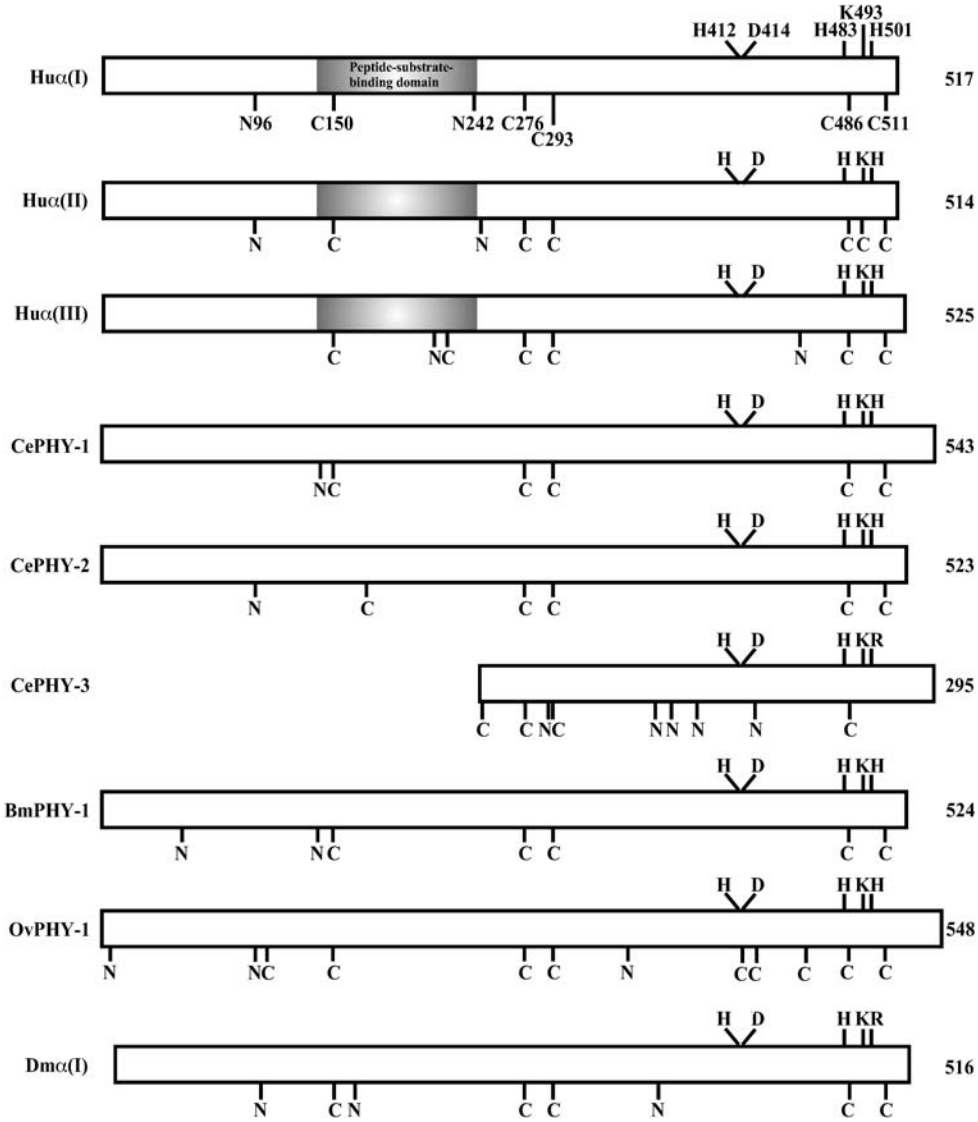

Fig. 3 Schematic representation of the human collagen prolyl 4-hydroxylase α(I), α(II) and α(III) subunits, the *C. elegans* PHY-1, PHY-2 and PHY-3 polypeptides, the *B. malayi* and *O. volvulus* PHY-1 polypeptides, and the *D. melanogaster* α(I) subunit. Numbering of the amino acids starts with the first residue of the processed polypeptide. Cysteine residues and potential attachment sites for asparagine-linked oligosaccharides are shown below the polypeptides, and the catalytically critical residues are shown above them. The peptide-substrate-binding domains in the human α subunits are also indicated

The human α(I), α(II) and α(III) subunits have five conserved cysteine residues, the α(II) and α(III) subunits each having one additional cysteine between the conserved cysteines 4 and 5, and 1 and 2, respectively [40, 44, 45] (Fig. 3). The collagen prolyl 4-hydroxylase tetramer has no interchain disulfide bonds between the subunits [52], but site-directed mutagenesis studies on the α(I) subunit have indicated that intrachain disulfide bonds that are essential for enzyme tetramer assembly are formed between the second and third conserved cysteines and between the fourth and fifth (Fig. 3) [53, 54]. The human α(I), α(II) and α(III) subunits have two *N*-glycosylation sites, the position of the more *N*-terminal site being conserved between the α(I) and α(II) subunits, while the other positions are not conserved (Fig. 3) [44, 45]. Site-directed mutagenesis studies of the human α(I) subunit and deglycosylation studies of the type III enzyme have shown that glycosylation has no role in the assembly of the enzyme tetramer or its catalytic activity [45, 54].

The genes encoding the human α(I), α(II) and α(III) subunits are present on chromosomes 10q21.3–23.1, 5q31 and 11q12, respectively, and have very similar exon-intron organizations [45, 55–57]. Despite this similarity, two forms of α(I) and α(II) mRNA have been described, resulting from mutually exclusive alternative splicing that affects different homologous exons, nos. 9 and 10 in the case of the α(I) gene, and 12a and 12b in that of the α(II) gene, whereas no evidence has been found for alternative splicing of the α(III) transcript [40, 45, 56, 57].

No human heritable diseases have been identified so far that are caused by mutations in the collagen prolyl 4-hydroxylase α subunit genes. Knock-out mice have recently been generated for the α(I) (Holster T, Pakkanen O, Soininen R, Sormunen R, Kivirikko KI, Myllyharju J, unpublished data) and α(II) subunits (Pakkanen O, Holster T, Soininen R, Sormunen R, Kivirikko KI, Myllyharju J, unpublished data). α(I) Knock-out causes embryonic lethality, whereas α(II) knock-out mice are born with no obvious phenotypic abnormalities.

3.1.1.2
The Multifunctional β Subunit

Cloning of the β subunit of human collagen prolyl 4-hydroxylase showed, surprisingly, that it is identical to another enzyme, protein disulfide isomerase (PDI) [58], an abundant protein within the ER that catalyzes disulfide bond formation and rearrangement during protein folding [5–7, 59]. PDI is a multifunctional polypeptide which serves as the β subunit in all currently known vertebrate collagen prolyl 4-hydroxylases and in the microsomal triglyceride transfer protein dimer, and as a chaperone-like polypeptide that binds various newly synthesized polypeptides within the lumen of the ER and probably assists in their folding [5–7, 59].

The human PDI polypeptide consists of 491 residues and a signal peptide of 17 additional residues [58]. It is a modular protein composed of four domains, *a* , *b* , *b′* and *a′*, and a highly acidic extension, *c* (Fig. 4) [5–7, 58, 59]. The *a* and

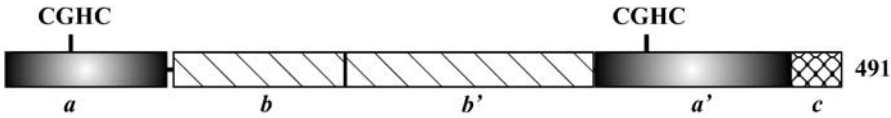

Fig. 4 Schematic representation of the domain structure of the human PDI polypeptide. Numbering of the amino acids starts with the first residue of the processed polypeptide. The two -Cys-Gly-His-Cys- sequences representing the catalytic sites for PDI activity are indicated. Modified from [5] with permission from Elsevier

a′ domains each contain a catalytic site for PDI activity, with a -Cys-Gly-His-Cys- motif, and are similar in sequence to thioredoxin [5–7, 58–60], while the homologous *b* and *b′* domains do not have catalytic site sequences and show no sequence similarity to thioredoxin [5–7, 58, 59]. NMR characterization of domains *a*, *b*, and *a′* indicates that they all have a thioredoxin fold, however [61–64]. PDI thus most probably consists of two catalytically active thioredoxin modules and two inactive ones [62].

PDI has at least two major functions in vertebrate collagen prolyl 4-hydroxylase tetramers. Its C terminus contains the -Lys-Asp-Glu-Leu motif, which is both necessary and sufficient for retention of the polypeptide, and hence also the collagen prolyl 4-hydroxylase tetramer, within the lumen of the ER [65]. When an enzyme tetramer is dissociated by various means [5, 6], or when the vertebrate α subunit is expressed alone in foreign host cells without PDI, the α subunit forms insoluble aggregates with no prolyl 4-hydroxylase activity [32, 33, 66]. Thus, PDI has an important function in keeping the α subunits in solution. Another chaperone, BiP, also forms soluble complexes with the α subunit, but with no prolyl 4-hydroxylase activity [67, 68]. Therefore, the function of PDI in the collagen prolyl 4-hydroxylase tetramer is not only that of keeping the α subunits in solution but is more specific, most likely that of keeping them in a catalytically active, non-aggregated conformation. Site-directed mutagenesis studies have shown that the disulfide isomerase activity of PDI is not required for the assembly or activity of the collagen prolyl 4-hydroxylase tetramer [65]. Likewise, the C-terminal extension *c* is not required for tetramer assembly [69], the minimum requirement being fulfilled by a pair of PDI domains, *b′a′* [70].

Besides its critical role as the β subunit of collagen prolyl 4-hydroxylase, PDI has additional important roles in collagen synthesis. It catalyzes intrachain disulfide bond formation in the N and C propeptides of procollagens and stabilizes procollagen trimers through the formation of interchain disulfide bonds between the C propeptides, and between the C telopeptides in some specific cases [1–4]. Furthermore, PDI interacts specifically with the C propeptides prior to trimer formation, retains unassembled chains within the ER, and thus directly prevents secretion until the correct triple-helical structure has been achieved [71, 72].

3.1.2
Non-Vertebrate Collagen Prolyl 4-Hydroxylases

Two major collagen families are present in the nematode *Caenorhabditis elegans*, the cuticle collagens and basement membrane collagens. These are encoded by a large gene family of about 180 members, three of them encoding basement membrane collagens and the rest cuticle collagens [4, 73, 74]. The *C. elegans* genome contains two conserved genes, *phy-1* and *phy-2*, encoding polypeptides with a high sequence similarity to the vertebrate collagen prolyl 4-hydroxylase α subunits, while the products of two additional genes, encoding α subunit-like polypeptides, show a lower similarity [75]. The *phy-1* and *phy-2* genes, and one with lower similarity, *phy-3*, have been cloned and characterized [76–81], while characterization of the other gene with lower similarity, *phy-4*, is in progress (Keskiaho K, Kukkola L, Page AP, Winter AD, Nissi R, Myllyharju J, unpublished data). PDI has two isoforms in *C. elegans*, PDI-1 and PDI-2, both having disulfide isomerase activity and PDI-2 also serving as the β subunit in the forms of collagen prolyl 4-hydroxylase that are involved in the synthesis of cuticle collagens [74, 77, 80, 82]. The PHY-1, PHY-2 and PDI-2 polypeptides are expressed in the cuticle collagen-synthesizing hypodermal cells at times of maximal collagen synthesis in larval stages 1–4 and in adult nematodes [77, 80].

The *C. elegans* PHY-1 and PHY-2 polypeptides consist of 543 and 523 residues, respectively, with signal peptides of 16 additional residues (Fig. 3) [76–78]. The degree of amino acid sequence identity between PHY-1 and PHY-2 is 54%, while that between PHY-1 and PHY-2 and the human α(I) and α(II) subunits is 42–46% [76, 77]. Surprisingly, recombinant expression of the *C. elegans* PHY-1 polypeptide in insect cells together with human PDI or *C. elegans* PDI-2 led to the assembly of an active αβ dimer instead of an $\alpha_2\beta_2$ tetramer (Fig. 2) [76, 82]. Recombinant expression and in vivo studies showed, however, that the main collagen prolyl 4-hydroxylase form in *C. elegans* is a unique PHY-1/PHY-2/$(\text{PDI-2})_2$ mixed tetramer, along with a small amount of the PHY-1/PDI-2 dimer, while the PHY-2/PDI-2 dimer was not detected (Fig. 2) [80]. Homozygous inactivation of either the *phy-1* or *phy-2* gene prevents assembly of the mixed tetramer, but the mutants can in part compensate for its absence by increased assembly of the corresponding PHY/PDI-2 dimer [80]. The *phy-1* mutants do this only very ineffectively, however, and have a short, fat phenotype, dumpy, whereas the *phy-2* mutants have a wild-type phenotype [80]. The *phy-1*$^{-/-}$, *phy-2*$^{-/-}$ double null [77, 78, 80] and the *pdi-2*$^{-/-}$ null [77] mutants lack all the collagen prolyl 4-hydroxylase forms needed for the synthesis of cuticle collagens and are therefore embryonically lethal.

The third *C. elegans* α subunit homologue characterized, PHY-3, encodes a polypeptide of only 295 residues, with a signal peptide of 23 additional residues (Fig. 3) [81]. It shows a 17–20% amino acid sequence identity to the about 290 C-terminal residues in the vertebrate α subunits and the *C. elegans* PHY-1 and PHY-2 polypeptides [81]. The recombinant PHY-3 polypeptide does not associate with *C. elegans* PDI-2, but instead has been shown in recombinant coex-

pression experiments to form an active collagen prolyl 4-hydroxylase with *C. elegans* PDI-1 that does not associate with PHY-1 or PHY-2 [80, 81]. The molecular composition of the active enzyme formed by PHY-3 and PDI-1 is still unknown, however, and it is also possible that PHY-3 may be a monomeric enzyme requiring PDI-1 only as a chaperone to assist in its folding (Fig. 2) [81]. Homozygous *phy-3* null nematodes are fertile and have a wild-type phenotype with a normal 4-hydroxyproline content in the cuticle, but the 4-hydroxyproline content of early mutant embryos was reduced by approximately 90% [81]. The *phy-3* gene is expressed in embryos, late larval stages and adult nematodes, its expression in adults being restricted to the spermatheca, a specialized region of the gonad where oocytes are fertilized [81]. PHY-3 therefore has no role in the synthesis of cuticle collagens but is likely to be involved in the hydroxylation of proline residues in early embryos, probably those in the eggshell collagens [81].

Collagen prolyl 4-hydroxylase α subunits have also been cloned and characterized from two parasitic filarial nematodes, *Onchocerca volvulus* and *Brugia malayi* (Fig. 3), the enzyme in the latter being a highly unusual α_4 homotetramer that is soluble and active without PDI (Fig. 2) [83, 84]. Small-molecule inhibitors of collagen prolyl 4-hydroxylase (see below) have similar cuticle-specific deleterious effects in both *C. elegans* and *B. malayi*, and thus this enzyme is an excellent target for the control of parasitic nematodes by chemical inhibition [78, 80, 83].

The collagen gene family of *Drosophila melanogaster* contains only four members, coding for two α chains of type IV collagen, one homologue of type XV and XVIII collagens and pericardin, a protein in which the collagen domain shows some similarity to type IV [4, 85–87]. It is therefore highly surprising that approximately 20 genes encoding polypeptides of 480–550 residues with a similarity to the vertebrate collagen prolyl 4-hydroxylase α subunits exist in the *D. melanogaster* genome [88]. Only one of the encoded polypeptides has been characterized in detail [89], and this consists of 516 residues with a signal peptide of an additional 19 residues and shows 34–35% and 31% amino acid sequence identities to the vertebrate α subunits and *C. elegans* PHY-1, respectively (Fig. 3) [89]. This *D. melanogaster* α subunit assembles into an active collagen prolyl 4-hydroxylase tetramer with PDI in recombinant expression (Fig. 2) [89]. Many of the *D. melanogaster* genes encoding α subunit-like polypeptides show tissue-specific and developmental stage-specific expression [88, 89]. The *D. melanogaster* collagen prolyl 4-hydroxylase family appears to be markedly larger than the corresponding vertebrate and *C. elegans* families, even though the number of collagen genes in *D. melanogaster* is much smaller. It may therefore be that many of the *D. melanogaster* prolyl 4-hydroxylases hydroxylate proline residues in proteins other than the collagens, and detailed studies on these enzymes may lead to the identification of additional functions for the human prolyl 4-hydroxylases as well.

3.2
Lysyl Hydroxylase

Lysyl hydroxylase has been purified to homogeneity from chick embryos and human placental tissues by two affinity column procedures [6, 31]. The active enzyme is a homodimer with a molecular weight of about 170 kDa, consisting of two identical monomers with a molecular weight of about 85 kDa. Lysyl hydroxylase is located in the lumen of the ER, but unlike the collagen prolyl 4-hydroxylases, it is not a soluble ER luminal protein but a luminally-oriented peripheral membrane protein [6]. A fully active recombinant lysyl hydroxylase can be efficiently produced in insect cells, and the use of a baculoviral signal peptide leads to efficient secretion of the enzyme into the culture medium, from which it can be easily purified [90–92].

3.2.1
Vertebrate Lysyl Hydroxylases

Vertebrate lysyl hydroxylase was first cloned from chicken and later from human, rat and mouse sources [93–97]. Many early findings suggested the possible existence of tissue-specific lysyl hydroxylase isoenzymes, e.g. the large differences in the extent of lysine hydroxylation found between collagen types, and even within the same collagen type from different tissues, and the wide variation in collagen hydroxylysine deficiency in different tissues of patients with the kyphoscoliotic type of Ehlers-Danlos syndrome (see below). Like collagen prolyl 4-hydroxylases, lysyl hydroxylase is now also known to form an enzyme family, since two novel isoenzymes, lysyl hydroxylases 2 and 3, have recently been cloned and characterized [98–100].

The human lysyl hydroxylase 1, 2 and 3 polypeptides consist of 709, 712 and 714 residues, respectively, with signal peptides of additional 18, 25 and 24 residues (Fig. 5) [94, 95, 98–100]. The overall amino acid sequence identity between the processed human lysyl hydroxylase 1 and 2 polypeptides is 75%, and that between lysyl hydroxylase 3 and the other two isoenzymes 57–59% [98–100]. Identity is highest within the catalytically important C-terminal region, being over 90% between lysyl hydroxylases 1 and 2, and 69–72% between lysyl hydroxylase 3 and the other two, all four catalytically critical residues being conserved (Fig. 5, see below) [98–100]. The vertebrate collagen prolyl 4-hydroxylase α subunits and lysyl hydroxylases show no significant amino acid sequence similarity with the exception of the catalytically critical residues. Further variation within the lysyl hydroxylase family is caused by alternative splicing. A novel form of human lysyl hydroxylase 2, termed lysyl hydroxylase 2b, is a 733-residue polypeptide, the increase in size being caused by sequences encoded by an additional exon, 13A, that is located between exons 13 and 14 in the originally described lysyl hydroxylase 2a cDNA [101].

The three human lysyl hydroxylase isoenzymes contain several conserved cysteine residues that may be structurally important (Fig. 5) [98–100]. All three

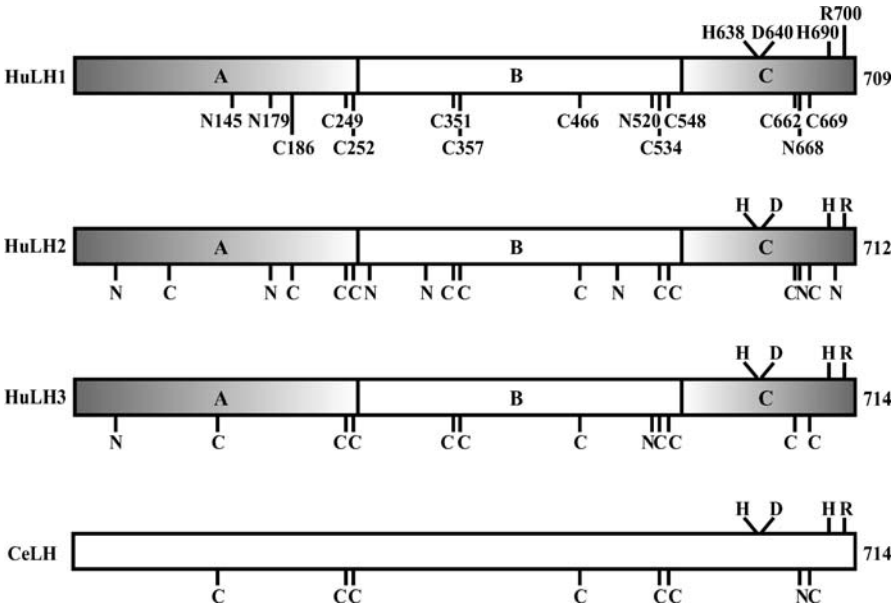

Fig. 5 Schematic representation of the human lysyl hydroxylase 1, 2 and 3 and *C. elegans* lysyl hydroxylase polypeptides. Numbering of the amino acids starts with the first residue of the processed polypeptide. Cysteine residues and potential attachment sites for asparagine-linked oligosaccharides are shown below the polypeptides, and the catalytically critical residues are shown above them. The domain structures of the human lysyl hydroxylase polypeptides are indicated by A, B and C

have potential asparagine-linked glycosylation sites, isoenzymes 1, 2 and 3 having four, seven and two sites, respectively (Fig. 5) [94, 95, 98–100]. A considerable heterogeneity is found in the glycosylation of human lysyl hydroxylase 1 [6], and site-directed mutagenesis studies have shown that some of the asparagine-linked carbohydrate units in a recombinant polypeptide are required for maximal enzyme activity [90]. Although the lysyl hydroxylases reside within the lumen of the ER, they do not contain a traditional ER retention motif, and it has been suggested that they are bound to the ER membrane by weak electrostatic interactions [102]. The 40 C-terminal residues of the lysyl hydroxylase 1 polypeptide have been shown to be necessary for its membrane association and localization in the ER [103, 104].

Limited proteolysis experiments on recombinant lysyl hydroxylase isoenzymes have indicated that the polypeptides consist of three domains A-C (from the N to C terminus), with molecular weights of approximately 30, 37 and 16 kDa (Fig. 5) [10]. The N-terminal domain A has no role in lysyl hydroxylase activity, as a recombinant B-C polypeptide was found to represent a fully active lysyl hydroxylase with catalytic properties identical to those of the full-length enzyme [10]. Lysyl hydroxylase 3, but not the other two isoenzymes, has also been reported recently to have small amounts of collagen glucosyltransferase

and galactosyltransferase activity [9–11], these activities residing entirely in the N-terminal domain A [10].

The human genes encoding lysyl hydroxylases 1, 2 and 3 are located on chromosomes 1p36.2–1p36.3, 3q23-q24 and 7q36, respectively [94, 100, 105]. Their mRNAs are expressed in a variety of human tissues, the highest expression levels for lysyl hydroxylase 1 mRNA being found in the liver and skeletal muscle (106), those for lysyl hydroxylase 2 mRNA in the pancreas, placenta, heart and skeletal muscle (98), and those for lysyl hydroxylase 3 in the pancreas, placenta, heart and spinal cord [99, 100]. No clear differences in tissue specificity or collagen type specificity have so far been identified between the three isoenzymes, but two mutations in the lysyl hydroxylase 2 gene have recently been reported in the autosomally recessive Bruck syndrome, which involves underhydroxylation of lysine residues within the telopeptides of type I collagen in bone and is characterized by fragile bones, joint contractures, scoliosis and osteoporosis [107]. This suggests that lysyl hydroxylase 2 may be a telopeptide lysyl hydroxylase with a tissue-specific expression pattern. Furthermore, it has been shown very recently that mice lacking lysyl hydroxylase 3 activity die during embryogenesis due to a lack of type IV collagen in their basement membranes [108]. This phenotype resembles that seen in *C. elegans* when its only lysyl hydroxylase is inactivated (see below).

Ehlers-Danlos syndrome is a heterogeneous group of heritable connective tissue disorders characterized by articular hypermobility, skin extensibility and tissue fragility [109]. The kyphoscoliotic type of Ehlers-Danlos syndrome, characterized by scoliosis, generalized joint laxity, skin fragility, ocular manifestations and severe muscle hypotonia, is caused by mutations in the lysyl hydroxylase 1 gene [109, 110]. The most common of these mutations leads to duplication of seven exons due to an Alu-Alu recombination in introns 9 and 16, which contain many Alu repeats [106, 111, 112]. The introns of the lysyl hydroxylase 3 gene are shorter than those of the lysyl hydroxylase 1 gene, but they also contain numerous Alu repeats [113].

3.2.2 *Caenorhabditis elegans* Lysyl Hydroxylase

The nematode *C. elegans* has a single gene encoding lysyl hydroxylase, *let -268* [99, 114]. The encoded polypeptide consists of 714 residues and a signal peptide of 16 additional residues [99, 114]. Its sequence shows 45% overall identity to those of the human lysyl hydroxylases 1–3, the identity being highest, 65%, within the 100 C-terminal residues [99, 114]. Inactivation of the *let -268* gene leads to the retention of type IV collagen in the producing cells, which indicates that lysyl hydroxylase activity is required for proper processing and secretion of type IV collagen [114]. The resulting lack of type IV collagen in the *C. elegans* basement membranes leads to separation of the muscle cells from the underlying epidermal layer upon body wall muscle contraction and failure to complete embryogenesis [114].

3.3
Prolyl 3-Hydroxylases

Prolyl 3-hydroxylase was purified to 5000-fold and partially characterized from a chick embryo extract already more than 25 years ago and its molecular weight was estimated to be about 160 kDa [6, 17]. However, prolyl 3-hydroxylase was cloned from chicken only very recently and it was found to be a chick homolog of leprecan, also known as growth suppressor 1 [8]. Leprecan, which is now called prolyl 3-hydroxylase 1, has two additional family members, prolyl 3-hydroxylases 2 and 3, previously called MLAT4 and GRBC in human, mouse and chicken [8]. The lengths of the processed human prolyl 3-hydroxylase 1, 2 and 3 polypeptides are 736, 708 and 736 amino acids, respectively, each having a C-terminal ER retention signal -Lys-Asp-Glu-Leu [8]. The sequence identity between the human prolyl 3-hydroxylase 1 and 2 polypeptides is 46% and that between the prolyl 3-hydroxylase 2 and 3 polypeptides 38%. The critical iron and 2-oxoglutarate binding residues are all conserved (see below), the 2-oxoglutarate binding residue being an arginine. Prolyl 3-hydroxylase 1 is localized specifically to tissues that express fibril-forming collagens, suggesting that it may serve to modify these collagens, while some of the other two isoenzymes may be involved in modifying the type IV basement membrane collagen [8].

3.4
Peptide Substrates of Collagen Hydroxylases

The collagen hydroxylases act on proline or lysine residues only in peptide linkages, and none of them hydroxylates the corresponding free amino acid. Collagen prolyl 4-hydroxylase does not act on tripeptides with the structure Gly-X-Pro or Pro-Gly-X in vitro, whereas X-Pro-Gly tripeptides are hydroxylated [5, 6, 17]. Likewise, lysyl hydroxylase does not act on a Lys-Gly-Pro tripeptide, while the X-Lys-Gly tripeptides are hydroxylated [6, 17]. Studies with various peptides have shown that the minimum sequence requirements for collagen prolyl 4-hydroxylase and lysyl hydroxylase are -X-Pro-Gly- and -X-Lys-Gly-, respectively [6, 17]. In addition, collagen prolyl 4-hydroxylase can hydroxylate the tetrapeptides Pro-Pro-Ala-Pro and Pro-Pro-Glu-Pro at a low rate, agreeing with the presence of a few -X-4Hyp-Ala- sequences in some collagen polypeptide chains and in the subcomponent C1q of complement [6, 17]. Lysyl hydroxylase can also hydroxylate the -X-Lys-Ala- and -X-Lys-Ser- sequences found in the telopeptides of fibril-forming collagens [6, 17]. Thus, the glycine in the -X-Y-Gly- sequence can in some rare cases be replaced by other amino acids. The interaction of collagen prolyl 4-hydroxylase and lysyl hydroxylase is also affected by the amino acid in the X position of the triplet to be hydroxylated and by other nearby amino acids [6, 17]. The chain length of the peptide has a major effect on the K_m values, which decrease with increasing chain length (Table 1) [5–7, 17]. The conformation of the peptide substrate has a crucial effect on

Table 1 K_m and K_i values of human type I–III collagen prolyl 4-hydroxylases (P4H-I-III) for $(Pro\text{-}Pro\text{-}Gly)_n$ and poly(L-proline), respectively

Substrate or inhibitor	Constant	P4H-I	P4H-II	P4H-III
	K_m or K_i, µmol/l			
$(Pro\text{-}Pro\text{-}Gly)_5$	K_m	150–250 [a]	490 [b]	N.D. [c]
$(Pro\text{-}Pro\text{-}Gly)_{10}$	K_m	18 [d]	95 [d]	20 [e]
Protocollagen[f]	K_m	0.2 [d]	1.1 [d]	N.D. [c]
Poly(L-proline), M_r 5000–7000	K_i	0.5 [d]	95 [d]	30 [e]
Poly(L-proline), M_r 44,000	K_i	0.02 [d]	20 [d]	N.D. [c]

[a] [31]; [b] [51]; [c]Not determined; [d] [44]; [e] [45]; [f]A biologically prepared substrate consisting of nonhydroxylated procollagen chains of chick type I procollagen. Ref. [31].

hydroxylation, in that the triple-helical conformation completely prevents proline and lysine hydroxylation [5–7, 17].

Some distinct differences in peptide-binding properties have been observed between the three vertebrate collagen prolyl 4-hydroxylase isoenzymes. The K_m values of the type I and type III enzymes for the peptide substrate $(Pro\text{-}Pro\text{-}Gly)_{10}$ are quite similar, while that of the type II enzyme is about fivefold higher (Table 1) [43–45]. Poly(L-proline) is a highly effective competitive inhibitor of the vertebrate type I collagen prolyl 4-hydroxylase, the K_i decreasing with increasing chain length (Table 1) [5–7, 17]. In contrast, the type II enzyme is inhibited by poly(L-proline) only at a much higher concentration, while the type III enzyme has intermediate inhibitory properties with respect to poly(L-proline) [43–45]. These findings suggest that distinct differences in the structures of the peptide-binding sites must exist between the three collagen prolyl 4-hydroxylase isoenzymes.

The peptide-substrate-binding domain in the collagen prolyl 4-hydroxylases is distinct from the catalytic domain and is located between residues Phe144 and Ser244 in the human α(I) subunit (Fig. 3) [50]. Site-directed mutagenesis studies have shown that most, although not all, of the differences in peptide binding between the type I and type II collagen prolyl 4-hydroxylases can be attributed to the presence of a glutamate and glutamine in the α(II) subunit positions corresponding to Ile182 and Tyr233 in α(I) [50]. The K_d values determined for the binding of several synthetic peptides to the recombinant α(I) and α(II) peptide-substrate-binding domains are very similar to the K_m and K_i values of these peptides as substrates and inhibitors of the type I and type II enzyme tetramers [51]. The K_d value of the α(I) peptide-substrate-binding domain for a hydroxylated peptide was found to be much higher than that for a non-hydroxylated one, indicating a marked decrease in the affinity of hydroxylated peptides for the domain [51]. It is thus evident that many characteristic features of the binding of peptides to the vertebrate collagen prolyl 4-hydroxylase isoenzymes can be explained by the properties of binding to the peptide-

substrate-binding domain rather than the catalytic domain [51]. Whether a similar domain exists in the lysyl hydroxylases remains to be established.

The peptide-substrate-binding domain of the collagen prolyl 4-hydroxylase α(I) subunit shows no amino acid sequence similarity to those of other proline-rich peptide-binding domains or profilin [50, 115–120]. It consists of five α helices and one short putative β strand, and thus its secondary structure is also distinct from the known structures of other proline-rich peptide-binding domains and profilin, which consist mainly of β strands [51, 115–120]. The other proline-rich peptide-binding modules typically bind their ligand via critical aromatic residues located on a hydrophobic path [115–120]. NMR analyses also suggest that the α(I) peptide-substrate-binding domain seems to have similar characteristic features, since binding of $(Pro\text{-}Pro\text{-}Gly)_2$ to this domain caused chemical shift changes mainly in hydrophobic residues [51]. The crystal structure of the α(I) peptide-substrate-binding domain has recently been solved and it has been found to consist of five α helices and belong to a family of teratricopeptide repeat domains that are involved in many protein–protein interactions [121]. The peptide substrates and the competitive inhibitor poly(L-proline) are suggested as becoming bound to a groove lined by tyrosines, the side-chains of which have a repeat distance similar to that of a poly(L-proline) type II helix [121].

Prolyl 3-hydroxylase hydroxylates -Pro-4Hyp-Gly- sequences but not -Pro-Pro-Gly- sequences [6, 17]. This is in agreement with the existence of 3-hydroxyproline in collagens only in the sequence -Gly-3Hyp-4Hyp-Gly- [6, 17]. As in the case of collagen prolyl 4-hydroxylase and lysyl hydroxylase, proline 3-hydroxylation is also affected by the amino acids in the nearby triplets, and by the length and conformation of the substrate [6, 17].

3.5 Cosubstrates and Reaction Mechanisms

Collagen prolyl 4-hydroxylase, prolyl 3-hydroxylase and lysyl hydroxylase belong to the group of 2-oxoglutarate dioxygenases, of which collagen prolyl 4-hydroxylase was one of the earliest to be discovered, is most extensively studied, and has often been regarded as a model for investigating other enzymes in the group [5–7]. All three collagen hydroxylases require Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate (Fig. 6), and their K_m values for these cosubstrates are very similar (Table 2).

The 2-oxoglutarate is stoichiometrically decarboxylated during hydroxylation, with one atom of the O_2 molecule being incorporated into the succinate and the other into the hydroxy group formed on the proline or lysine residue (Fig. 6) [5–7, 17]. Ascorbate is not consumed stoichiometrically, and the enzymes can complete a number of reaction cycles at an essentially maximal rate in its absence [5–7, 17]. Hydroxylation then ceases, however, and ascorbate is required to reactivate the enzyme [5, 6, 17]. The reaction requiring ascorbate is an uncoupled decarboxylation of 2-oxoglutarate, i.e., decarboxylation with-

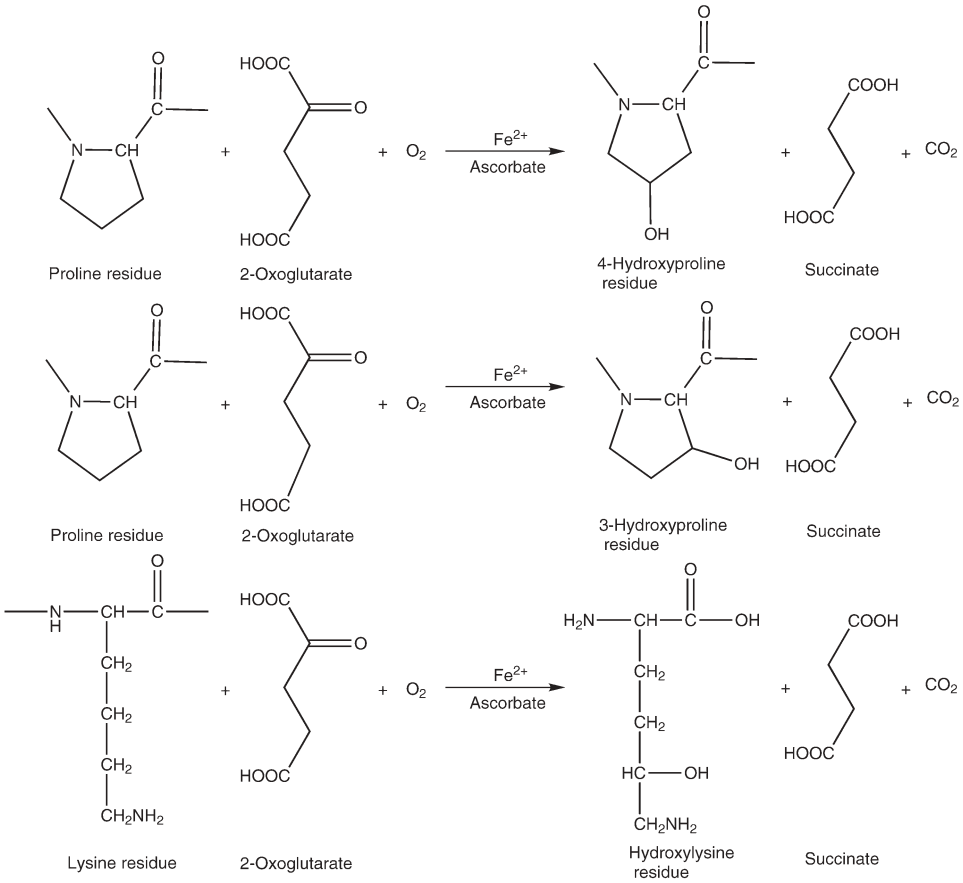

Fig. 6 Reactions catalyzed by collagen prolyl 4-hydroxylase, prolyl 3-hydroxylase and lysyl hydroxylase

Table 2 K_m values of human type I–III collagen prolyl 4-hydroxylases (P4H-I-III), human lysyl hydroxylases 1 and 3 (LH1 and 3) and chick prolyl 3-hydroxylase (P3H) for cosubstrates. The values for collagen prolyl 4-hydroxylases and lysyl hydroxylases have been determined with synthetic peptides $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Ile\text{-}Lys\text{-}Gly)_3$ as substrates, respectively, whereas those for prolyl 3-hydroxylase have been determined with a protocollagen substrate, the values for this enzyme being therefore slightly lower

Cosubstrate	P4H-I	P4H-II	P4H-III	LH1	LH3	P3H[a]
	K_m, µmol/l					
Fe^{2+}	2 [a]	2 [b]	0.5 [c]	2 [d]	2 [d]	2 [e]
2-Oxoglutarate	20 [a]	22 [f]	20 [c]	100 [d]	100 [d]	3 [e]
Ascorbate	300 [a]	340 [f]	370 [c]	350 [d]	300 [d]	120 [e]
O_2	40 [g]	N.D. [h]	N.D. [h]	45 [e,i]	N.D. [h]	30 [e]

[a] [49]; [b] [56]; [c] [45]; [d] [99]; [e] [17]; [f] [44]; [g] [30]; [h]Not determined; [i]Chick LH1.

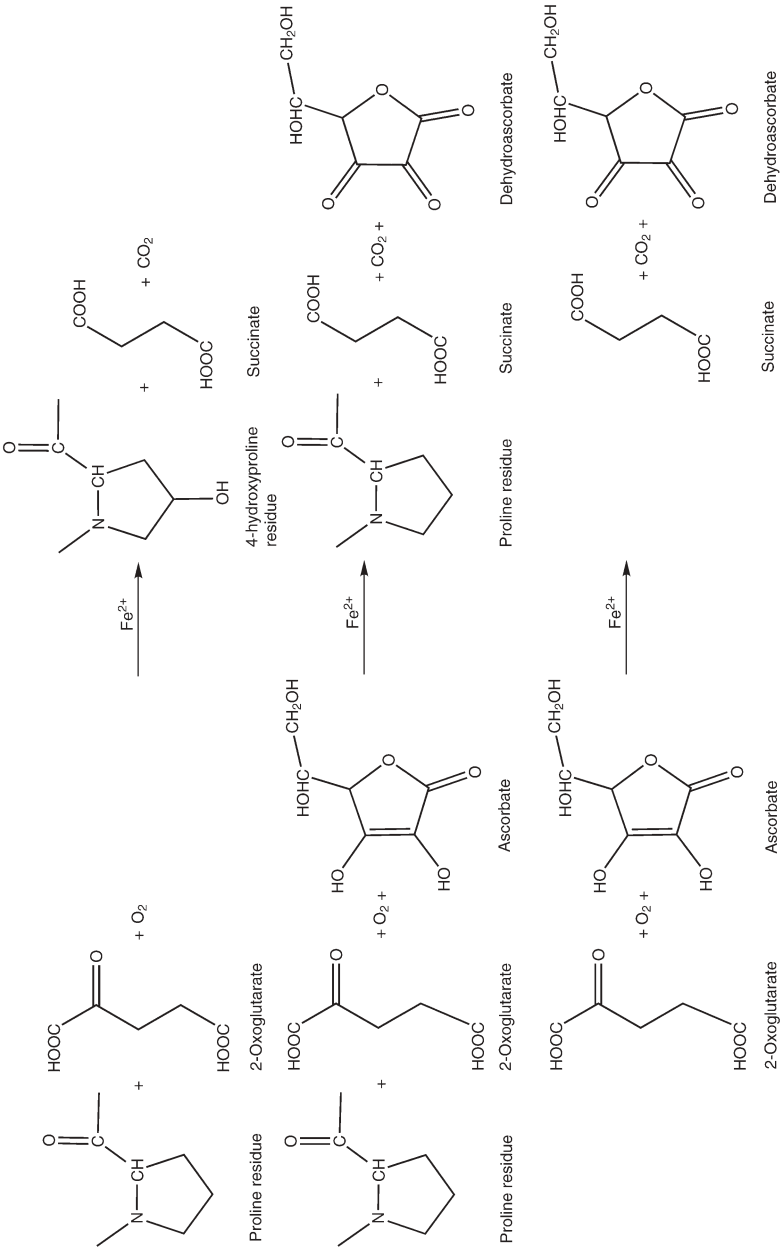

Fig. 7 Reactions catalyzed by collagen prolyl 4-hydroxylase. 2-Oxoglutarate is stoichiometrically decarboxylated during the hydroxylation reaction, which does not require ascorbate (**A**). The enzyme also catalyzes the uncoupled decarboxylation of 2-oxoglutarate without subsequent hydroxylation of the peptide substrate. Ascorbate is stoichiometrically consumed in the uncoupled reaction, which may occur in either the presence (**B**) or absence (**C**) of the peptide substrate

out subsequent hydroxylation, in which ascorbate is consumed stoichiometrically (Fig. 7). The collagen hydroxylases catalyze uncoupled decarboxylation cycles even in the presence of saturating concentrations of the peptide substrates, and in these cycles the reactive iron-oxo complex is probably converted to $Fe^{3+}{\cdot}O^{-}$ rendering the enzymes unavailable for new catalytic cycles until reduced by ascorbate [5–7, 17]. The main biological function of ascorbate in the collagen hydroxylase reactions is therefore probably to serve as an alternative oxygen acceptor in the uncoupled decarboxylation cycles [5–7, 17].

Kinetic studies of the collagen prolyl 4-hydroxylase and lysyl hydroxylase reactions have shown that the cosubstrates and the peptide substrate become bound to the enzymes in an ordered manner, Fe^{2+} binding first, followed by 2-oxoglutarate, O_2 and the peptide substrate, and the reaction products are released in the reverse order, although Fe^{2+} is not released between most catalytic cycles [6, 17]. The catalytic cycle can be divided into two half-reactions: initial generation of the reactive hydroxylating species and its subsequent utilization for hydroxylation of the target residue (Fig. 8) [5–7, 17].

The current stereochemical model for a catalytic site in the collagen hydroxylases suggests that it consists of a set of separate locations for binding of the cosubstrates [5–7, 17, 122, 123]. The Fe^{2+} is located in a pocket coordinated with three side chains (Fig. 8). Site-directed mutagenesis studies and other data have indicated that His412, Asp414, and His483 are the Fe^{2+} binding residues in the human α(I) subunit (Figs. 3 and 8), while the corresponding ligands in the lysyl hydroxylase 1 polypeptide are His638, Asp640 and His690 (Fig. 5) [49, 54, 90, 124]. The 2-oxoglutarate binding site can be divided into three subsites [122, 123, 125]: subsite I is a positively charged residue of the enzyme that ionically binds the C5 carboxyl group of the 2-oxoglutarate, subsite II consists of two *cis*-positioned coordination sites of the enzyme-bound Fe^{2+} and is chelated by the C1-C2 moiety, while subsite III involves a hydrophobic binding site in the C3-C4 region of 2-oxoglutarate (Fig. 8) [122, 123, 125]. Site-directed mutagenesis studies indicate that subsite I is formed by Lys493 in the human collagen prolyl 4-hydroxylase α(I) subunit and by Arg700 in the corresponding position in human lysyl hydroxylase 1 (Figs. 3 and 5) [49, 126]. The three Fe^{2+} binding residues are conserved in all the collagen prolyl 4-hydroxylase α subunit isoforms and lysyl hydroxylase isoenzymes characterized so far (Figs. 3 and 5), as well as in other 2-oxoglutarate dioxygenases characterized, and a related enzyme, isopenicillin N synthase [27–29, 127–135]. The positively charged residue forming subsite I of the 2-oxoglutarate binding site is likewise conserved in position +9 or +10 with respect to the second Fe^{2+} binding histidine in all 2-oxoglutarate dioxygenases with the exception of HIF asparaginyl hydroxylase in which it is located in position +15 with respect to the first Fe^{2+} binding histidine [27–29, 127–135]. Determination of the crystal structures of several 2-oxoglutarate dioxygenases and isopenicillin N synthase has verified the critical role of these conserved residues in Fe^{2+} and 2-oxoglutarate binding [127–135]. Although the overall amino acid sequence similarity between the enzymes is very low, their catalytic sites are all located in a jelly-roll motif formed by eight

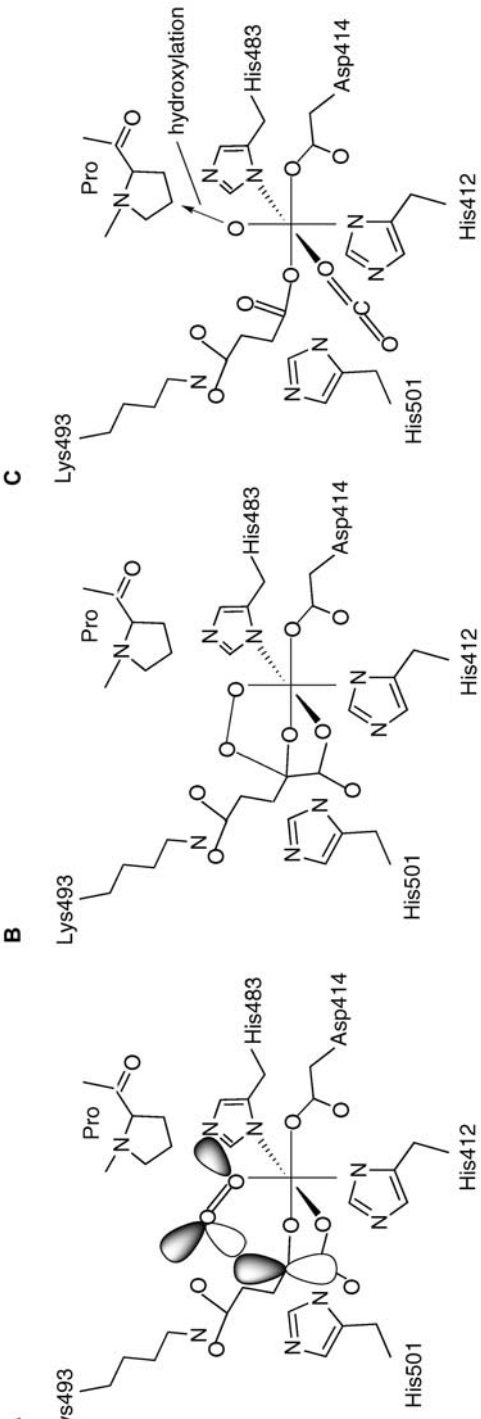

Fig. 8 Schematic representation of the first half-reaction of collagen prolyl 4-hydroxylase and the critical residues at the catalytic site of the human α(I) subunit. Fe^{2+} is coordinated with the enzyme by His412, Asp414 and His483. Subsite I of the 2-oxoglutarate binding site consists of Lys493, which ionically binds the C5 carboxyl group of 2-oxoglutarate, while subsite II consists of two *cis* -positioned equatorial coordination sites of the enzyme-bound Fe^{2+} and is chelated by the C1 carboxyl and C2 oxo functions of the 2-oxoglutarate. Molecular oxygen is bound end-on in an axial position, producing a dioxygen unit. (**A**) One of the electron-rich orbitals of the dioxygen is directed to the electron-depleted orbital at C2 of the 2-oxoglutarate bound to the iron. (**B**) A nucleophilic attack on C2 generates a tetrahedral intermediate, with loss of the double bond in the dioxygen unit and of the double-bond characteristics of the oxo-acid moiety. (**C**) Elimination of CO_2 coincides with the formation of succinate and a ferryl ion, which hydroxylates a proline residue in the peptide substrate in the second half-reaction. His501 is an additional important residue, probably being involved in both coordination of the C1 carboxyl group of 2-oxoglutarate with Fe^{2+} and cleavage of the tetrahedral ferryl intermediate. Modified from [5] with permission from Elsevier

β strands. It is thus probable that a similar jelly-roll core can also be found at the catalytic sites of the collagen hydroxylases.

The His501 residue in the human collagen prolyl 4-hydroxylase α(I) subunit (Fig. 3) is an additional critical residue [49, 54], which probably has two roles at the catalytic site: it directs the orientation of the C1 carboxyl group of 2-oxoglutarate to the active iron centre, and it accelerates the breakdown of the tetrahedral ferryl intermediate to succinate, CO_2 and a ferryl ion (Fig. 8B) [49]. The ascorbate binding site of the collagen hydroxylases probably also contains the two *cis*-positioned coordination sites of the enzyme-bound iron discussed above, and is thus partially identical to subsite II of the 2-oxoglutarate binding site [136].

Molecular oxygen is thought to become bound to the Fe^{2+} end-on in an axial position, producing the dioxygen unit, a species characterized by doubly occupied orbitals (Fig. 8) [122, 123]. One of the electron-rich orbitals of the dioxygen unit is directed to the electron-depleted orbital at C2 of the 2-oxoglutarate bound to the iron atom (Fig. 8A), so that the C2 undergoes rehybridization from its sp^2 hybridized planar oxo structure to an sp^3 hybridized tetrahedral transition state and forms a covalent bond with the non-coordinated atom of the dioxygen unit (Fig. 8B). This weakens both the C-C bond in the 2-oxoglutarate and the O-O bond in the dioxygen unit [122, 123]. Decarboxylation then occurs concurrently with cleavage of the O-O bond, and the original C2 of 2-oxoglutarate, which has become the C1 of succinate, returns to the sp^2 hybridization. Simultaneously, a highly reactive ferryl ion is formed (Fig. 8B), which hydroxylates the proline or lysine residue in the peptide substrate in the second half-reaction [122, 123]. This second half-reaction renews the enzyme-bound Fe^{2+} and concludes the catalytic cycle.

3.6 Inhibitors and Inactivators

The formation of scars and fibrous tissue is part of the normal healing process after injury, but in some situations collagen accumulates in excessive amounts, leading to fibrosis, which compromises the normal functioning of the affected tissue. The central role of collagen in fibrosis has prompted attempts to develop drugs that inhibit its accumulation, and the critical function of 4-hydroxyproline in collagen has made collagen prolyl 4-hydroxylase an attractive target for antifibrotic therapy. Many compounds that inhibit collagen prolyl 4-hydroxylase, and in many cases the other two collagen hydroxylases as well, are now known, although no such inhibitors are yet in clinical use.

Collagen hydroxylase inhibitors with respect to their peptide substrates and all cosubstrates are now known. Many peptides are competitive inhibitors with respect to the peptide substrate, poly (L-proline), for example, being highly effective for the type I collagen prolyl 4-hydroxylase [5–7, 17], whereas its inhibitory potency with regard to the vertebrate type II and type III enzymes (Table 1) (see above) and the *C. elegans* and *D. melanogaster* collagen prolyl 4-

hydroxylases is much weaker [43–45, 76, 80, 89]. Many bivalent cations are competitive inhibitors with respect to Fe^{2+} [5, 6, 17], the most potent being Zn^{2+} with a K_i of 1 µmol/l for collagen prolyl 4-hydroxylase and lysyl hydroxylase [137, 138]. Superoxide dismutase-active copper chelates, including Cu(acetylsalicylate)$_2$ and Cu(lysine)$_2$, are competitive inhibitors with respect to O_2, the K_i of collagen prolyl 4-hydroxylase and lysyl hydroxylase for both compounds being 30 µmol/l [139].

A number of aliphatic and aromatic structural analogues of 2-oxoglutarate are competitive inhibitors with respect to this cosubstrate (Fig. 9). The most potent of these, pyridine 2,4-dicarboxylate and pyridine 2,5-dicarboxylate, have functional groups that can interact at all three subsites of the 2-oxoglutarate binding site, their K_i values for collagen prolyl 4-hydroxylase being 2 µmol/l and 0.8 µmol/l, respectively [136, 140, 141]. In the cases of prolyl 3-hydroxylase and lysyl hydroxylase, pyridine 2,4-dicarboxylate, with K_i values of 3 µmol/l and 50 µmol/l, respectively, is a more potent inhibitor than pyridine 2,5-dicarboxylate [141]. Lysyl hydroxylase has higher K_i values for almost all 2-oxoglutarate analogues than the two prolyl hydroxylases [141], this being in accordance with the higher K_m of lysyl hydroxylase for 2-oxoglutarate (Table 2). These and other data (see previous section) indicate that the three collagen hydroxylases have similar but not identical 2-oxoglutarate binding sites. Systematic variation of the structures of aromatic 2-oxoglutarate analogues has indicated that the K_i increases markedly if the iron chelating moiety is destroyed or the carboxyl group that becomes bound at subsite I (see previous section) is omitted or shifted to a position in which it cannot become bound at this subsite [140]. Several 2-oxoglutarate analogues also inhibit the other class of prolyl 4-hydroxylases, HIF prolyl 4-hydroxylases, but their inhibitory properties differ distinctly from those of the collagen prolyl 4-hydroxylases [30]. It should thus be possible to develop potent small molecule inhibitors that show a high degree of specificity with respect to the two classes of prolyl 4-hydroxylases, enabling selective targeting of fibrotic or ischaemic diseases.

N-Oxalylglycine (Fig. 9) is also a potent collagen prolyl 4-hydroxylase inhibitor with a K_i of about 0.5–8 µmol/l [142, 143]. It differs from 2-oxoglutarate only by replacement of the methylene group at C3 with -NH-, and is a competitive inhibitor with respect to 2-oxoglutarate, but cannot replace it as a cosubstrate [142, 143]. This is consistent with the reaction mechanism (see previous section), in which decarboxylation involves a nucleophilic attack by an electron pair in the dioxygen unit on the electron-deficient orbital at C2 of 2-oxoglutarate (Fig. 8). The C3 in 2-oxoglutarate is sp^3 hybridized and cannot contribute any electron density to the p_z orbital of C2 [122, 123, 143]. In oxalylglycine the C3 is replaced by the sp^2 hybridized nitrogen atom, the p_z orbital of which has an electron pair that can contribute to the electron density of C2, and therefore a nucleophilic attack is no longer possible [143]. Oxalylalanine, which differs from oxalylglycine by carrying a methyl group at the carbon atom that corresponds to the C4 of 2-oxoglutarate, is also an effective collagen pro-

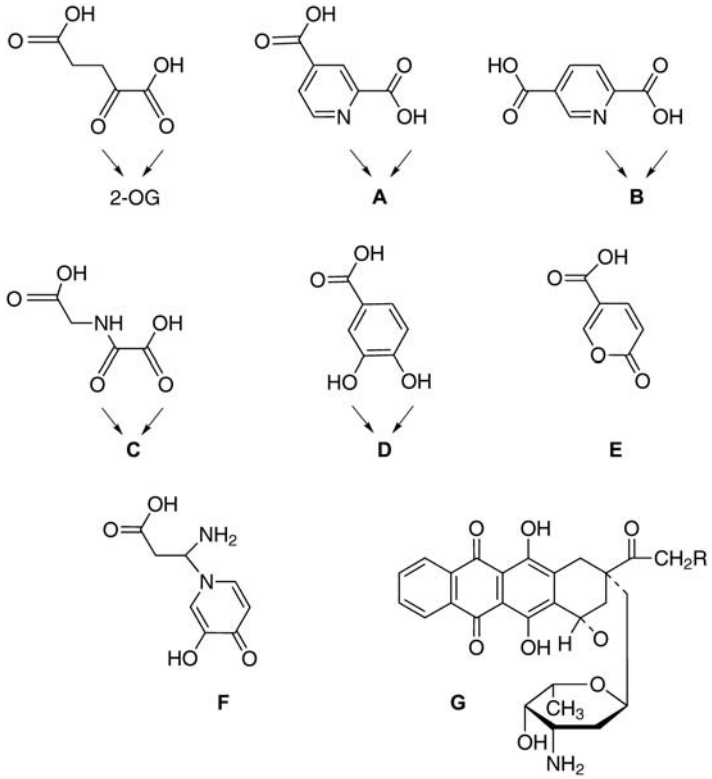

Fig. 9 Structures of some inhibitors of collagen hydroxylases. (2-OG) 2-oxoglutarate, (**A**) pyridine 2,4-dicarboxylate, (**B**) pyridine 2,5-dicarboxylate, (**C**) *N*-oxalylglycine, (**D**) 3,4-dihydroxybenzoate, (**E**) L-mimosine, (**F**) coumalic acid and (**G**) doxorubicin. Modified from [5] with permission from Elsevier

lyl 4-hydroxylase inhibitor, with a K_i of 40 µmol/l, while the other oxalyl amino acid derivatives show little inhibitory activity [142, 143].

3,4-Dihydroxybenzoate (Fig. 9) also possesses the functional groups required for interaction at all three subsites of the 2-oxoglutarate binding site and has a K_i of about 5 µmol/l for collagen prolyl 4-hydroxylase [136]. This compound and its derivatives differ from the pyridine derivatives in that they are competitive with respect to both 2-oxoglutarate and ascorbate, whereas the latter are uncompetitive with respect to ascorbate [136]. This conforms to the current stereochemical model for the catalytic site, according to which the ascorbate binding site is partially identical to the 2-oxoglutarate binding site [122, 123]. The dihydroxybenzoate and pyridine inhibitors probably become bound to different enzyme forms, as determined by the oxidation state of the iron atom at the catalytic site [136].

L-Mimosine (Fig. 9) inhibits collagen prolyl 4-hydroxylase by 50% at a 120 µmol/l concentration in addition to causing reversible, dose-dependent in-

hibition of DNA synthesis due to inhibition of the hydroxylation of deoxyhypusine to hypusine, a rare amino acid required by the eukaryotic initiation factor eIF-5A and for cell cycle transition [144]. This compound thus inhibits both collagen prolyl 4-hydroxylase and deoxyhypusyl hydroxylase, leading to both fibrosuppressive and antiproliferative effects [144].

Many of the known collagen hydroxylase inhibitors have low membrane permeability and are therefore ineffective in cultured cells and in vivo. The generation of lipophilic proinhibitor derivatives of these compounds has been used to overcome this problem, however. The proinhibitors do not themselves act as inhibitors of the pure enzymes, but readily pass through cell membranes and become processed to the active inhibitor only within the cell. Such proinhibitors include ethylpyridine 2,4-dicarboxylate [145], pyridine 2,4-dicarboxylic acid di(methoxyethyl)amide (HOE 077) [146], dimethyloxalylglycine [143] and ethyl 3,4-dihydroxybenzoate [147, 148].

Three groups of compounds have also been identified that act as irreversible inactivators of collagen prolyl 4-hydroxylase, probably by a suicide mechanism. The first group consists of peptides in which the proline residue to be hydroxylated is replaced by 5-oxaproline, a proline analogue containing oxygen as part of the five-membered ring [149]. The most potent of these peptides studied has the structure benzyloxycarbonyl-Phe-Oxaproline-Gly-benzylester and inactivates collagen prolyl 4-hydroxylase by 50% in 1 h at 0.8 µmol/l concentration [149]. Oxaproline-containing peptides also inactivate the enzyme in cultured human skin fibroblasts, although a similar degree of inactivation is obtained only at a concentration one order of magnitude higher than that required with the pure enzyme [150]. The second group includes coumalic acid (Fig. 9), which acts as a 2-oxoglutarate analogue but is only a weak inactivator, due to the lack of a functional group needed to become bound to the Fe^{2+} at a catalytic site, a 2 mmol/l concentration being required for 50% inactivation in 1 h [151]. The third group consists of the anthracyclines doxorubicin and daunorubicin (Fig. 9), which inactivate collagen prolyl 4-hydroxylase and lysyl hydroxylase by 50% in 1 h at 60 µmol/l and 150 µmol/l concentrations, respectively [152]. This inactivation can be prevented by high concentrations of ascorbate or low concentrations of its competitive analogues, whereas 2-oxoglutarate and its competitive analogues offer no protection [152].

More recently identified inhibitors include the lithospermic acid magnesium salt isolated from *Salviae miltorrhizae Radix*, a Chinese medicinal herb [153], the organophosphate insecticides malathion and malaoxon [154], and a heterocyclic carbonyl-glycine derivative S4682, which inhibits collagen prolyl 4-hydroxylase with a particularly low K_i of 155 nmol/l [155]. Minoxidil, an antihypertensive piperidinopyrimidine nitrooxide is unique among the collagen hydroxylase inhibitors in that it does not inhibit the enzyme reaction but specifically reduces the mRNA level of lysyl hydroxylase 1, and concurrently also the amount of the enzyme, whereas the amount of mRNAs for type I collagen prolyl 4-hydroxylase are unchanged [156–158].

4
Collagen Glycosyltransferases

Collagens contain carbohydrate units, either the monosaccharide galactose or the disaccharide glucosylgalactose, linked to hydroxylysine residues [6, 17]. The extent of glycosylation of hydroxylysine residues and the ratio of galactosylhydroxylysine to glucosylgalactosylhydroxylysine vary markedly between collagen types and within the same collagen type in various physiological and pathological states [159]. The functions of the hydroxylysine-linked carbohydrate units are not fully understood, but in the case of the fibril-forming collagens they influence the lateral packing of the collagen molecules into fibrils and the diameters of these fibrils both in vivo and in vitro [160, 161].

The formation of the hydroxylysine linked carbohydrate units is catalyzed by two specific enzymes, hydroxylysyl galactosyltransferase and galactosylhydroxylysyl glucosyltransferase (Fig. 10) [159]. The former has been partially

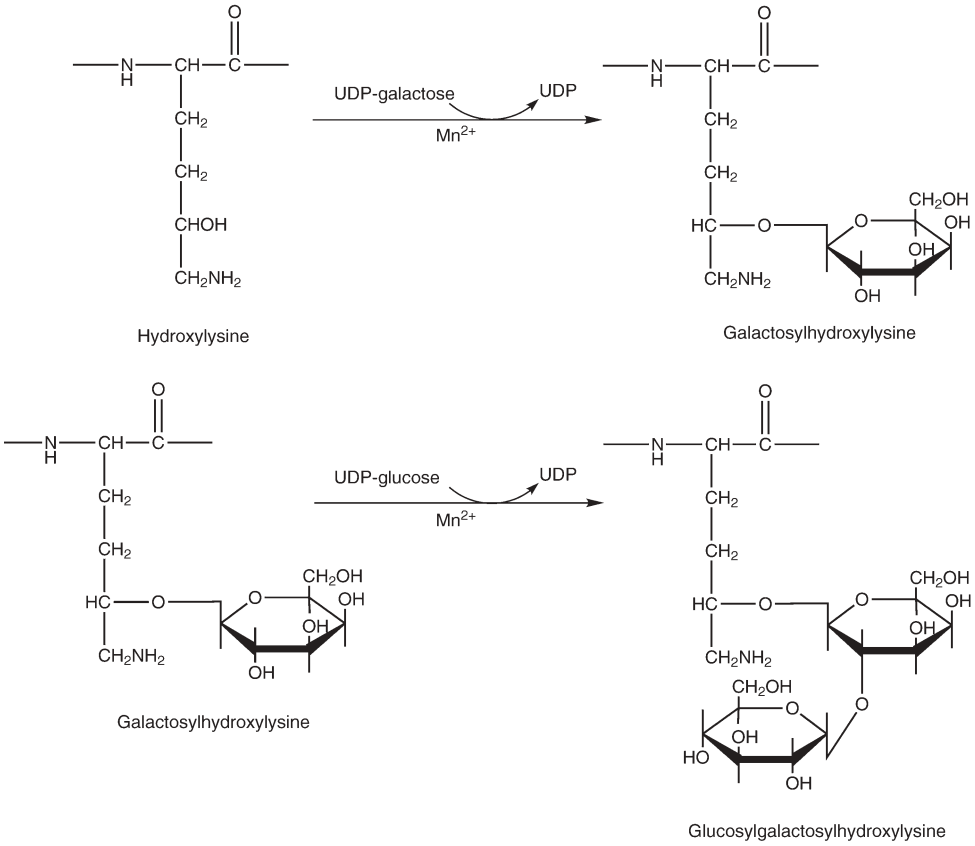

Fig. 10 Reactions catalyzed by hydroxylysyl galactosyltransferase (**A**) and galactosylhydroxylysyl glucosyltransferase (**B**)

purified from various sources, the activity being found in gel filtration in three forms with molecular weights of about 450 and 200 kDa and a minor species of about 50 kDa [162], while the latter has been purified to homogeneity and found to have a molecular weight of 72–78 kDa in SDS-PAGE analysis, but a lower molecular weight in gel filtration, probably on account of retarded elution due to partial adsorption into the column material [163]. These two enzymes have not been cloned yet, but interestingly, human lysyl hydroxylase 3 and the single *C. elegans* lysyl hydroxylase have been found to possess both collagen galactosyltransferase and glucosyltransferase activities [9–11, 164], although their levels in the human lysyl hydroxylase 3 polypeptide are so low that they may be of little biological significance. Furthermore, previous data suggesting that the two glycosyltransferase reactions may be catalyzed by separate enzymes in vertebrates [31, 162, 163] supports the existence of additional collagen glycosyltransferases that are likely to be responsible for most of the collagen glycosylation that takes place in vivo. The collagen glycosyltransferase activities of lysyl hydroxylase 3 reside in the *N*-terminal domain A, which has no role in the hydroxylase activity of the enzyme (Fig. 5) [10]. Site-directed mutagenesis studies have identified Cys144 and Leu208 in this human polypeptide as important for its glucosyltransferase activity and Cys144 and Asp187–191 as being important for galactosyltransferase activity, whereas none of these residues plays a role in hydroxylase activity [164].

Collagen galactosyltransferase catalyzes the transfer of galactose from UDP-galactose to hydroxylysine residues, while collagen glucosyltransferase catalyzes the transfer of glucose from UDP-glucose to galactosylhydroxylysine residues (Fig. 10) [159]. The free ε-amino group of the hydroxylysine residue is an absolute requirement for both reactions, which are also influenced by the amino acid sequence of the peptide, the peptide chain length and the peptide conformation, the triple-helical conformation preventing both glycosylations [159]. Both reactions require a bivalent cation, preferably Mn^{2+} [159].

Acknowledgments Laura Silvennoinen is acknowledged for preparing Figs. 6, 7 and 10 and for modifying Figs. 8 and 9.

References

1. Prockop DJ, Kivirikko KI (1995) Annu Rev Biochem 64:403
2. Lamandé SR, Bateman JF (1999) Semin Cell Dev Biol 10:455
3. Kielty CM, Grant ME (2002) The collagen family: structure, assembly, and organization in the extracellular matrix. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders. Molecular, genetic and medical aspects, 2nd edn. Wiley-Liss, New York, p 159
4. Myllyharju J, Kivirikko KI (2004) Trends Genet 20:33

5. Kivirikko KI, Myllyharju J (1998) Matrix Biol 16:357
6. Kivirikko KI, Pihlajaniemi T (1998) Adv Enzymol Related Areas Mol Biol 72:325
7. Myllyharju J (2003) Matrix Biol 22:15
8. Vranka JA, Sakai LY, Bächinger HP (2004) J Biol Chem 279:23615
9. Heikkinen J, Risteli M, Wang C, Latvala J, Rossi M, Valtavaara M, Myllylä R (2000) J Biol Chem 275:36158
10. Rautavuoma K, Takaluoma K, Passoja K, Pirskanen A, Kvist A-P, Kivirikko KI, Myllyharju J (2002) J Biol Chem 277:23084
11. Wang C, Luosujärvi H, Heikkinen J, Risteli M, Uitto L, Myllylä R (2002) Matrix Biol 21:559
12. Jimenez S, Harsch M, Rosenbloom J (1973) Biochem Biophys Res Commun 52:106
13. Berg RA, Prockop DJ (1973) Biochem Biophys Res Commun 52:115
14. Jenkins CL, Raines RT (2002) Nat Prod Rep 19:49
15. Jenkins CL, Bretscher LE, Guzei IA, Raines RT (2003) J Am Chem Soc 125:6422
16. Mizuno K, Hayashi T, Peyton DH, Bächinger HP (2004) J Biol Chem 279:282
17. Kivirikko KI, Myllylä R, Pihlajaniemi T (1992) Hydroxylation of proline and lysine residues in collagens and other animal proteins. In: Harding JJ, Crabbe MJC (eds) Post-translational modifications of proteins. CRC Press, Boca Raton, p 1
18. Maeda H, Matsumura Y, Kato H (1988) J Biol Chem 263:16051
19. Sasaguri M, Ikeda M, Ideishi M, Arakawa K (1988) Biochem Biophys Res Commun 150:511
20. Gautron J-P, Pattou E, Bauer K, Kordon C (1991) Neurochem Int 18:221
21. Andrews PC, Hawke D, Shively JE, Dixon JE (1984) J Biol Chem 259:15021
22. Spiess J, Noe BD (1985) Proc Natl Acad Sci USA 82:277
23. Wijffels GL, Panaccio M, Salvatore L, Wilson L, Walker ID, Spithill TW (1994) Biochem J 299:781
24. Bozas SE, Spithill TW (1996) Exp Parasitol 82:69
25. Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, Salic A, Asara JM, Lane WS, Kaelin WG Jr (2001) Science 292:464
26. Jaakkola P, Mole DR, Tian Y-M, Wilson MI, Gielbert J, Gaskell SJ, von Kriegsheim A, Hebestreit HF, Mukherji M, Schofield CJ, Maxwell PH, Pugh CW, Ratcliffe PJ (2001) Science 292:468
27. Bruick RK, McKnight SL (2001) Science 294:1337
28. Epstein ACR, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole RD, Mukherji M, Metzen E, Wilson MI, Dhanda A, Tian Y-M, Masson N, Hamilton DL, Jaakkola P, Barstead R, Hodgkin J, Maxwell PH, Pugh CW, Schofield CJ, Ratcliffe PJ (2001) Cell 107:43
29. Ivan M, Haberberger T, Gervasi DC, Michelson KS, Günzler V, Kondo K, Yang H, Sorokina I, Conaway RC, Conaway JW, Kaelin WG Jr (2002) Proc Natl Acad Sci USA 99:13459
30. Hirsilä M, Koivunen P, Günzler V, Kivirikko KI, Myllyharju J (2003) J Biol Chem 278:30772
31. Kivirikko KI, Myllylä R (1982) Methods Enzymol 82:245
32. Vuori K, Pihlajaniemi T, Marttila M, Kivirikko KI (1992) Proc Natl Acad Sci USA 89:7467
33. Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI (1997) EMBO J 16:6702
34. Merle C, Perret S, Lacour T, Jonval V, Hudaverdian S, Garrone R, Ruggiero F, Theisen M (2002) FEBS Lett 515:114
35. Lamberg A, Helaakoski T, Myllyharju J, Peltonen S, Notbohm H, Pihlajaniemi T, Kivirikko KI (1996) J Biol Chem 271:11988
36. Myllyharju J, Lamberg A, Notbohm H, Fietzek PP, Pihlajaniemi T, Kivirikko KI (1997) J Biol Chem 272:21824

37. Toman PD, Chisholm G, McMullin H, Giere LM, Olsen DR, Kovach RJ, Leigh SD, Fong BE, Chang R, Daniels GA, Berg RA, Hitzeman RA (2000) J Biol Chem 275:23303
38. Nokelainen M, Tu H, Vuorela A, Notbohm H, Kivirikko KI, Myllyharju J (2001) Yeast 18:1
39. Walmsley AR, Battern MR, Lad U, Bulleid NJ (1999) J Biol Chem 274:14884
40. Helaakoski T, Vuori K, Myllylä R, Kivirikko KI, Pihlajaniemi T (1989) Proc Natl Acad Sci USA 86:4392
41. Bassuk JA, Kao WW-Y, Herzer P, Kedersha NL, Seyer J, DeMartino JA, Daugherty BL, Mark GE III, Berg RA (1989) Proc Natl Acad Sci USA 86:7382
42. Hopkinson I, Smith SA, Donne A, Gregory H, Franklin TJ, Grant ME, Rosamund J (1994) Gene 149:391
43. Helaakoski T, Annunen P, Vuori K, MacNeil IA, Pihlajaniemi T, Kivirikko KI (1995) Proc Natl Acad Sci USA 92:4427
44. Annunen P, Helaakoski T, Myllyharju J, Veijola J, Pihlajaniemi T, Kivirikko KI (1997) J Biol Chem 272:17342
45. Kukkola L, Hieta R, Kivirikko KI, Myllyharju J (2003) J Biol Chem 278:47685
46. Van Den Diepstraten C, Papay K, Bolender Z, Brown A, Pickering JG (2003) Circulation 108:508
47. Annunen P, Autio-Harmainen H, Kivirikko KI (1998) J Biol Chem 273:5989
48. Nissi R, Autio-Harmainen H, Marttila P, Sormunen R, Kivirikko KI (2001) J Histochem Cytochem 49:1143
49. Myllyharju J, Kivirikko KI (1997) EMBO J 16:1173
50. Myllyharju J, Kivirikko KI (1999) EMBO J 18:306
51. Hieta R, Kukkola L, Permi P, Pirilä P, Kivirikko KI, Kilpeläinen I, Myllyharju J (2003) J Biol Chem 278:34966
52. Nietfeld JJ, Van Der Kraan J, Kemp A (1981) Biochim Biophys Acta 661:21
53. John DCA, Bulleid NJ (1994) Biochemistry 33:14018
54. Lamberg A, Pihlajaniemi T, Kivirikko KI (1995) J Biol Chem 270:9926
55. Pajunen L, Jones TA, Helaakoski T, Pihlajaniemi T, Sheer D, Solomon E, Kivirikko KI (1989) Am J Hum Genet 45:829
56. Nokelainen M, Nissi R, Kukkola L, Helaakoski T, Myllyharju J (2001) Eur J Biochem 268:5300
57. Helaakoski T, Veijola J, Vuori K, Rehn M, Chow LT, Taillon-Miller P, Kivirikko KI, Pihlajaniemi T (1994) J Biol Chem 269:27847
58. Pihlajaniemi T, Helaakoski T, Tasanen K, Myllylä R, Huhtala M-L, Koivu J, Kivirikko KI (1987) EMBO J 6:643
59. Freedman RB, Klappa P, Ruddock LW (2002) EMBO Rep 3:136
60. Vuori K, Myllylä R, Pihlajaniemi T, Kivirikko KI (1992) J Biol Chem 267:7211
61. Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE (1996) Biochemistry 35:7684
62. Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE (1997) Curr Biol 7:239
63. Kemmink J, Dijkstra K, Mariani M, Scheek RM, Penka E, Nilges M, Darby NJ (1999) J Biomol NMR 13:357
64. Dijkstra K, Karvonen P, Pirneskoski A, Koivunen P, Kivirikko KI, Darby NJ, van Straaten M, Scheek RM, Kemmink J (1999) J Biomol NMR 14:195
65. Vuori K, Pihlajaniemi T, Myllylä R, Kivirikko KI (1992) EMBO J 11:4213
66. John DCA, Grant ME, Bulleid NJ (1993) EMBO J 12:1587
67. Veijola J, Pihlajaniemi T, Kivirikko KI (1996) Biochem J 315:613
68. John DCA, Bulleid NJ (1996) Biochem J 317:659
69. Koivunen P, Pirneskoski A, Karvonen P, Ljung J, Helaakoski T, Notbohm H, Kivirikko KI (1999) EMBO J 18:65

70. Pirneskoski A, Ruddock LW, Klappa P, Freedman RB, Kivirikko KI, Koivunen P (2001) J Biol Chem 276:11287
71. Wilson R, Lees JF, Bulleid NJ (1998) J Biol Chem 273:9637
72. Bottomley MJ, Batten MR, Lumb RA, Bulleid NJ (2001) Curr Biol 11:1114
73. Johnstone IL (2000) Trends Genet 16:21
74. Page AP, Winter AD (2003) Adv Parasitol 53:85
75. The *C. elegans* Sequencing Consortium (1998) Science 282:2012
76. Veijola J, Koivunen P, Annunen P, Pihlajaniemi T, Kivirikko KI (1994) J Biol Chem 269:26746
77. Winter AD, Page AP (2000) Mol Cell Biol 20:4084
78. Friedman L, Higgin JJ, Moulder G, Barstead R, Raines RT, Kimble J (2000) Proc Natl Acad Sci USA 97:4736
79. Hill KL, Harfe BD, Dobbins CA, L'Hernault SW (2000) Genetics 155:1139
80. Myllyharju J, Kukkola L, Winter AD, Page AP (2002) J Biol Chem 277:29187
81. Riihimaa P, Nissi R, Page AP, Winter AD, Keskiaho K, Kivirikko KI, Myllyharju J (2002) J Biol Chem 277:18238
82. Veijola J, Annunen P, Koivunen P, Page AP, Pihlajaniemi T, Kivirikko KI (1996) Biochem J 317:721
83. Merriweather A, Günzler V, Brenner M, Unnasch TR (2001) Mol Biochem Parasitol 116:185
84. Winter AD, Myllyharju J, Page AP (2003) J Biol Chem 278:2554
85. Hynes RO, Zhao Q (2000) J Cell Biol 150:F89
86. Ackley BD, Crew JR, Elamaa H, Pihlajaniemi T, Kuo CJ, Kramer JM (2001) J Cell Biol 152:1219
87. Chartier A, Zaffran S, Astier M, Semeriva M, Gratecos D (2002) Development 129:3241
88. Abrams EW, Andrew DJ (2002) Mech Dev 112:165
89. Annunen P, Koivunen P, Kivirikko KI (1999) J Biol Chem 274:6790
90. Pirskanen A, Kaimio A-M, Myllylä R, Kivirikko KI (1996) J Biol Chem 271:9398
91. Krol BJ, Murad S, Walker LC, Marshall MK, Clark WL, Pinnell SR, Yeowell HN (1996) J Invest Dermatol 106:11
92. Rautavuoma K, Takaluoma K, Passoja K, Pirskanen A, Kvist A-P, Kivirikko KI, Myllyharju J (2002) J Biol Chem 277:23084
93. Myllylä R, Pihlajaniemi T, Pajunen L, Turpeenniemi-Hujanen T, Kivirikko KI (1991) J Biol Chem 266:2805
94. Hautala T, Byers MG, Eddy RL, Shows TB, Kivirikko KI, Myllylä R (1992) Genomics 13:62
95. Yeowell HN, Ha V, Walker LC, Murad S, Pinnell SR (1992) J Invest Dermatol 99:864
96. Armstrong LC, Last JA (1995) Biochim Biophys Acta 1264:93
97. Ruotsalainen H, Sipilä L, Kerkelä E, Pospiech H, Myllylä R (1999) Matrix Biol 18:325
98. Valtavaara M, Papponen H, Pirttilä AM, Hiltunen K, Helander H, Myllylä R (1997) J Biol Chem 272:6831
99. Passoja K, Rautavuoma K, Ala-Kokko L, Kosonen T, Kivirikko KI (1998) Proc Natl Acad Sci USA 95:10482
100. Valtavaara M, Szpirer C, Szpirer J, Myllylä R (1998) J Biol Chem 273:12881
101. Yeowell HN, Walker LC (1999) Matrix Biol 18:179
102. Kellokumpu S, Sormunen R, Heikkinen J, Myllylä R (1994) J Biol Chem 269:30524
103. Suokas M, Myllylä R, Kellokumpu S (2000) J Biol Chem 275:17863
104. Suokas M, Lampela O, Juffer AH, Myllylä R, Kellokumpu S (2003) Biochem J 370:913
105. Szpirer C, Szpirer J, Rivière M, Vanvooren P, Valtavaara M, Myllylä R (1997) Mamm Genome 8:707

106. Heikkinen J, Hautala T, Kivirikko KI, Myllylä R (1994) Genomics 24:464
107. Van der Slot AJ, Zuurmond A-M, Bardoel AFJ, Wijmenga C, Pruijs HEH, Sillence DO, Brinckmann J, Abraham DJ, Black CM, Verzijl N, DeGroot J, Hanemaaijer R, TeKoppele JM, Huizinga TWJ, Bank RA (2003) J Biol Chem 278:40967
108. Rautavuoma K, Takaluoma K, Sormunen R, Myllyharju J, Kivirikko KI, Soininen R (2004) Proc Natl Acad Sci USA 101:14120
109. Steinmann B, Royce PM, Superti-Furga A (2002) The Ehlers-Danlos syndrome. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders. Molecular, genetic and medical aspects, 2nd edn. Wiley-Liss, New York, p 431
110. Yeowell HN, Walker LC (2000) Mol Genet Metab 71:212
111. Hautala T, Heikkinen J, Kivirikko KI, Myllylä R (1993) Genomics 15:399
112. Heikkinen J, Toppinen T, Yeowell HN, Krieg T, Steinmann B, Kivirikko KI, Myllylä R (1997) Am J Hum Genet 60:48
113. Rautavuoma K, Passoja K, Helaakoski T, Kivirikko KI (2000) Matrix Biol 19:73
114. Norman KR, Moerman DG (2000) Dev Biol 227:690
115. Kay B, Williamson MP, Sudol M (2000) FASEB J 14:231
116. Macias MJ, Wiesner S, Sudol M (2002) FEBS Lett 513:30
117. Ball LJ, Jarchau T, Oschkinat H, Walter U (2002) FEBS Lett 513:45
118. Freund C, Dötsch V, Nishizawa K, Reinherz EL, Wagner G (1999) Nat Struct Biol 6:656
119. Mahoney NM, Rozwarski DA, Fedorov E, Fedorov AA, Almo SC (1999) Nat Struct Biol 6:666
120. Nodelman IM, Bowman GD, Lindberg U, Schutt C (1999) J Mol Biol 294:1271
121. Pekkala M, Hieta R,Bergmann U, Kivirikko KI, Wierenga RK, Myllyharju J (2004) J Biol Chem 279:52255
122. Hanauske-Abel HM, Günzler V (1982) J Theor Biol 94:421
123. Hanauske-Abel HM, Popowicz AM (2003) Curr Med Chem 10:1005
124. Myllylä R, Günzler V, Kivirikko KI, Kaska DD (1992) Biochem J 286:923
125. Majamaa K, Hanauske-Abel HM, Günzler V, Kivirikko KI (1984) Eur J Biochem 138:239
126. Passoja K, Myllyharju J, Pirskanen A, Kivirikko KI (1998) FEBS Lett 434:145
127. Roach PL, Clifton IJ, Fülöp V, Harlos K, Barton GJ, Hajdu J, Andersson I, Schofield CJ, Baldwin JE (1995) Nature 375:700
128. Valegård K, Terwisscha van Scheltinga AC, Lloyd MD, Hara T, Ramaswamy S, Perrakis A, Thompson A, Lee H-J, Baldwin JE, Schofield CJ, Hajdu J, Andersson I (1998) Nature 394:805
129. Clifton IJ, Hsueh L-C, Baldwin JE, Harlos K, Schofield CJ (2001) Eur J Biochem 268:6625
130. Elkins JM, Ryle MJ, Clifton IJ, Dunning Hotopp JC, Lloyd JS, Burzlaff NI, Baldwin JE, Hausinger RP, Roach PL (2002) Biochemistry 41:5185
131. Wilmouth RC, Turnbull JJ, Welford RWD, Clifton IJ, Prescott AG, Schofield CJ (2002) Structure 10:93
132. Dann CE III, Bruick RK, Deisenhofer J (2002) Proc Natl Acad Sci USA 99:15351
133. Elkins JM, Hewitson KS, McNeill LA, Seibel JF, Schlemminger I, Pugh CW, Ratcliffe PJ, Schofield CJ (2003) J Biol Chem 278:1802
134. Lee C, Kim SJ, Jeong DG, Lee SM, Ryu SE (2003) J Biol Chem 278:7558
135. Clifton IJ, Doan LX, Sleeman MC, Topf M, Suzuki H, Wilmouth RC, Schofield CJ (2003) J Biol Chem 278:20843
136. Majamaa K, Günzler V, Hanauske-Abel HM, Myllylä R, Kivirikko KI (1986) J Biol Chem 261:7819
137. Tuderman L, Myllylä R, Kivirikko KI (1977) Eur J Biochem 80:341
138. Puistola U, Turpeenniemi-Hujanen T, Myllylä R, Kivirikko KI (1980) Biochim Biophys Acta 611:51

139. Myllylä R, Schubotz LM, Weser U, Kivirikko KI (1979) Biochem Biophys Res Commun 89:98
140. Majamaa K, Hanauske-Abel HM, Günzler V, Kivirikko KI (1984) Eur J Biochem 138:239
141. Majamaa K, Turpeenniemi-Hujanen TM, Latipää P, Günzler V, Hanauske-Abel HM, Hassinen IE, Kivirikko KI (1985) Biochem J 229:127
142. Cunliffe CJ, Franklin TJ, Hales NJ, Hill GB (1992) J Med Chem 35:2652
143. Baader E, Tschank G, Baringhaus K-H, Burghard H, Günzler V (1994) Biochem J 300:525
144. McCaffrey TA, Pomerantz K, Sanborn TA, Spokojny A, Du B, Park MH, Folk JE, Lamberg A, Kivirikko KI, Falcone DJ, Mehta S, Hanauske-Abel HM (1995) J Clin Invest 95:446
145. Tschank G, Brocks DG, Engelbart K, Mohr J, Baader E, Günzler V, Hanauske-Abel HM (1991) Biochem J 275:469
146. Hanauske-Abel HM (1991) J Hepatol 13(Suppl 3):S8
147. Majamaa K, Sasaki T, Uitto J (1987) J Invest Dermatol 89:405
148. Sasaki T, Majamaa K, Uitto J (1987) J Biol Chem 262:9397
149. Günzler V, Brocks DG, Henke S, Myllylä R, Geiger R, Kivirikko KI (1988) J Biol Chem 263:19498
150. Karvonen K, Ala-Kokko L, Pihlajaniemi T, Helaakoski T, Günzler V, Henke S, Kivirikko KI, Savolainen E-R (1990) J Biol Chem 265:8415
151. Günzler V, Hanauske-Abel HM, Myllylä R, Mohr J, Kivirikko KI (1987) Biochem J 242:163
152. Günzler V, Hanauske-Abel HM, Myllylä R, Kaska DD, Hanauske A, Kivirikko KI (1988) Biochem J 251:365
153. Shigematsu T, Tajima S, Nishikawa T, Murad S, Pinnell SR, Nishioka I (1994) Biochim Biophys Acta 1200:79
154. Samimi A, Last JA (2001) Toxicol Appl Pharmocol 172:203
155. Bickel M, Baringhaus KH, Gerl M, Günzler V, Kanta J, Schmidts L, Stapf M, Tschank G, Weidmann K, Werner U (1998) Hepatology 28:404
156. Murad S, Pinnell SR (1987) J Biol Chem 262:11973
157. Hautala T, Heikkinen J, Kivirikko KI, Myllylä R (1992) Biochem J 283:51
158. Yeowell HN, Ha V, Walker LC, Murad S, Pinnell SR (1992) J Invest Dermatol 99:864
159. Kivirikko KI, Myllylä R (1979) Int Rev Connect Tissue Res 8:23
160. Brinckmann J, Notbohm H, Tronnier M, Acil Y, Fietzek PP, Schmeller W, Müller PK, Batge B (1999) J Invest Dermatol 113:617
161. Notbohm H, Nokelainen M, Myllyharju J, Fietzek PP, Müller PK, Kivirikko KI (1999) J Biol Chem 274:8988
162. Risteli L, Myllylä R, Kivirikko KI (1976) Biochem J 155:145
163. Myllylä R, Anttinen H, Risteli L, Kivirikko KI (1977) Biochim Biophys Acta 480:113
164. Wang C, Risteli M, Heikkinen J, Hussa A-K, Uitto L, Myllylä R (2002) J Biol Chem 277:18568

Top Curr Chem (2005) 247: 149–183
DOI 10.1007/b103822
© Springer-Verlag Berlin Heidelberg 2005

Biosynthetic Processing of Collagen Molecules

Daniel S. Greenspan (✉)

Department of Pathology and Laboratory Medicine, University of Wisconsin, Medical School, 1300 University Avenue MSC 5675, Madison, Wisconsin 53706, USA
dsgreens@wisc.edu

1 Introduction . . . 150

2 Processing of the Major Fibrillar Collagens . . . 151
2.1 N-Proteinases and the ADAMTS Family of Metalloproteinases . . . 152
2.2 C-Proteinases and the BMP-1/Tolloid Family of Metalloproteinases . . . 154
2.3 Major Procollagen Processing in the Context of Secretion and Fibrillogenesis . . . 157
2.4 C-Proteinase Enhancer Proteins . . . 158
2.5 Noncollagenous Substrates of the BMP-1/Tolloid-Related C-Proteinases . . . 159
2.5.1 Small Leucine-Rich Proteoglycans . . . 160
2.5.2 Lysyl Oxidase . . . 162
2.5.3 SIBLING Proteins . . . 162
2.5.4 Laminin-5 . . . 163
2.5.5 Myostatin/GDF-8 . . . 164

3 Processing of the Minor Fibrillar Collagens . . . 165

4 Processing of Type VII Collagen . . . 170

5 Processing/Shedding of Cell Surface Collagen Types XIII, XVII and XXV . . . 171

6 Processing of Multiplexin Collagen Types XV and XVIII . . . 173

7 Concluding Remarks . . . 175

References . . . 176

Abstract A large number of extracellular proteins are synthesized in vivo as precursor molecules that must be proteolytically processed into their mature functional forms. Included among such proteins are a subset of the collagens. A number of the proteinases involved in the biosynthetic processing of collagen precursor molecules have fairly recently been identified and characterized. It is now evident that the proteinases involved in processing collagen precursors are responsible for the biosynthetic processing of various other kinds of extracellular proteins as well. The result is coordination of biosynthesis of the various processed collagens, and the formation of macromolecular structures formed by some of them, with other molecular events in morphogenesis and homeostasis. In this chapter, the natures of recently characterized families of proteinases involved in collagen processing are summarized, as are recently described ancillary proteins that enhance collagen processing.

Also summarized are noncollagenous substrates that have recently been demonstrated to be processed by proteinases which also process collagens and whose biosyntheses are therefore linked with collagen biosynthesis.

Keywords Metalloproteinases · ADAMTS · Bone morphogenetic protein-1 · Tolloid · Proprotein convertases

List of Abbreviations

ECM	Extracellular matrix
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
BMP	Bone morphogenetic protein
C-propeptide	Carboxyl-terminal propeptide
dpc	Days post conception
CUB	Complement-Uegf-BMP-1
DEB	Dystrophic epidermolysis bullosa
DMP	Dental matrix protein
DSPP	Dentin stalophosphoprotein
ECM	Extracellular matrix
EDA	Ectodysplasin A
EDS	Ehlers-Danlos syndrome
GDF	Growth and differentiation factor
MMP	Matrix metalloproteinase
mTLD	Mammalian Tolloid
mTLL	Mammalian Tolloid-like
MEF	Mouse embryo fibroblast
NTR	Netrin
N-Propeptide	Amino-terminal propeptide
pCP	Procollagen C-proteinase
PCPE	Procollagen C-proteinase enhancer
pNP	Procollagen N-proteinase
SIBLING	Small integrin binding ligand N-linked glycoprotein
SLRP	Small leucine rich proteoglycans
TGF-β	Transforming growth factor β
TIMP	Tissue inhibitor of matrix metalloproteinase

1 Introduction

Proteolytic processing is a mechanism for the intracellular and extracellular regulation of protein function. Such processing can be catabolic, and act in the break down and removal of proteins, or anabolic, and represent important steps in the biosynthesis and functional activation of proteins. This chapter will deal with what is presently known concerning the anabolic aspects of collagen processing. These include those processing interactions involved in the biosynthesis and functional maturation of the fibrillar collagen types I–III, V and XI, and the nonfibrillar collagen type VII; and those processing interactions that appear to be involved in altering functions of the transmembrane collagens XIII, XVII and XXV, via shedding from the cell surface. Proteolytic processing

of the multiplexin collagen types XV and XVIII to yield anti-angiogenic fragments such as endostatin will be addressed, although it is not yet clear which of the multiple cleavages of these collagens may be involved in generating physiologically functional fragments, and which are involved in catabolic turnover. However, catabolic degradation of collagens by the matrix metalloproteinases (MMPs) will not be covered herein, and for this subject the reader is directed to several existing reviews [1–3].

This chapter will also deal with the natures of various noncollagenous substrates recently shown to be biosynthetically processed by the same proteinases that biosynthetically process the major fibrillar collagens, since processing of different substrates by the same proteinases suggests co-regulation of the in vivo events in which these different substrates participate. Thus, the nature of noncollagenous substrates cleaved by the same proteinases that process collagens provides insights into the types of in vivo molecular events with which collagen biosynthesis is coordinated.

2 Processing of the Major Fibrillar Collagens

Collagen types I–III constitute the major fibrous components of vertebrate extracellular matrix (ECM). All three of these major fibrillar collagens are synthesized as procollagen precursors that differ from mature collagen monomers in that they contain NH_2- and COOH-terminal peptide extensions, known as N- and C-propeptides, respectively (Fig. 1). These are cleaved, thus releasing the central triple helical portion of the molecule which serves as a mature monomer capable of associating into fibrils [4]. Type I procollagen is a heterotrimer comprising two pro-α1(I) chains and one pro-α2(I) chain. Mutations which remove the site for cleavage of the N-propeptide portion of the pro-α1(I) or pro-α2(I) chains, result in the heritable connective tissue disorders Ehlers-Danlos syndrome (EDS) types VIIA and VIIB, respectively [5, 6]. These are marked by joint hypermobility and multiple joint dislocations [7]. Deficiency in levels of the proteolytic activity responsible for cleaving both the pro-α1(I) and pro-α2(I) N-propeptide regions results in EDS type VIIC, marked by extreme fragility of the skin [7]. In EDS VIIB, there is little effect on fibril formation in dermis, whereas dermal fibrils are irregular in EDS VIIA and are even more so in EDS VIIC [7]. Thus, impairment of cleavage of the type I procollagen N-propeptide is not incompatible with in vivo fibrillogenesis, although these defects in N-propeptide processing can lead to abnormal fibril morphology in tissues. Consistent with what has been observed in vivo, studies with in vitro fibrillogenesis systems have shown that collagen monomers which retain the N-propeptide are readily incorporated into growing fibrils along with normal monomers, although inclusion of the abnormal monomers creates fibrils with abnormal morphologies [8]. In contrast, type I collagen monomers which retain the bulkier C-propeptide are not incorporated into growing fibrils

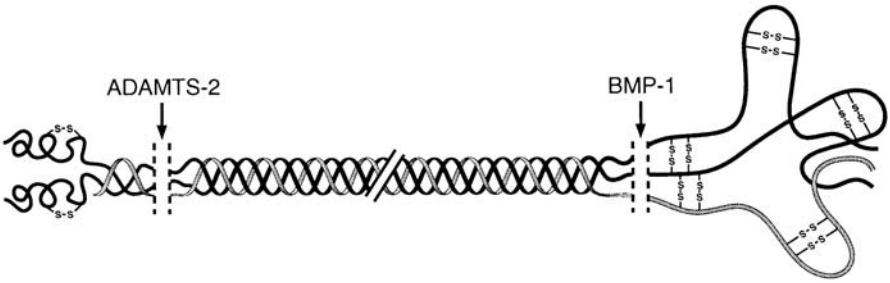

Fig. 1 Processing of type I procollagen N- and C-propeptides by ADAMTS-2 and BMP-1, respectively

in vitro [8], and no heritable disorders have yet been associated with inability to cleave the C-propeptide. Thus, it may be that inability to cleave major fibrillar procollagen C-propeptides is incompatible with fibrillogenesis, and therefore with morphogenesis and viability. The inhibitory effects of uncleaved C-propeptides on fibrillogenesis may be twofold: 1) via interfering with self-association of collagen monomers by markedly increasing their solubility; and 2) via steric hindrance of the highly ordered packing of monomers necessary for fibrillogenesis [4, 8].

2.1 N-Proteinases and the ADAMTS Family of Metalloproteinases

Accumulated biochemical evidence has shown both N- and C-propeptides of the major fibrillar procollagens to be cleaved by Ca^{2+}-dependent metalloproteinases with neutral pH optima and well defined spectra of proteinase inhibitors [9–13]. Cleavage of the N-propeptides of procollagen type I and the homotrimeric procollagen type II is via a procollagen type I N-proteinase (pNPI) activity that cleaves at a specific Pro-Gln site in pro-α1(I) chains and a specific Ala-Gln site in pro-α2(I) chains, and which has been well characterized in terms of modes of inhibition [9, 10, 14, 15]. Interestingly, this activity does not cleave heat denatured type I procollagen [15, 16], indicating a dependence on native conformation of the substrate. It has been reported that pNPI activity does not process the homotrimeric procollagen type III [9, 10, 16], the N-propeptide of which has been reported to be cleaved via a procollagen type III N-proteinase (pNPIII) activity that does not cleave procollagen type I [17–19]. Biochemical isolation of a protein with pNPI activity and partial amino acid sequencing [20] led to cloning and characterization of full-length cDNA sequences for both bovine and human forms of a pNP1 proteinase [21, 22]. Furthermore, mutations in the gene that encodes this pNPI were shown to underlie both the EDS VIIC phenotype in humans and the analogous bovine disease dermatosparaxis [22]. Sequence comparisons showed the pNP1 to belong to the ADAMTS (ADisintegrin And Metalloproteinase with ThromboSpondin motifs) family of met-

alloproteinases [23], of which there are presently 19 reported vertebrate family members [24, 25]. Since pNP1 was only the second ADAMTS proteinase to be identified, it has been designated ADAMTS-2. The ADAMTS proteinases share a common domain structure (Fig. 2) and resemble the ADAM (ADisintegrin And Metalloproteinase) family of proteinases in having pro-, adamalysin/reprolysin-like metalloprotease, disintegrin-like and cysteine-rich domains [23, 26]. However, they differ from many ADAM proteinases in lacking transmembrane domains and, instead, they possess variable numbers of thrombospondin type I-like repeats [23, 26] which, at least in some ADAMTS proteinases, appear to be involved in binding to ECM components [27]. Evidence indicates that the ADAMTS proteinases play a broad spectrum of roles in development, reproduction, disease and homeostasis. For example, ADAMTS-1, -4 and -5/11, which form a subset of ADAMTS proteinases based on degree of sequence homology and similarities in protein domain structure, cleave the proteoglycan aggrecan, and this aggrecanase activity is important both to the homeostasis of cartilage and to etiology of the arthritides [28–30]. ADAMTS-1 was originally identified as an inflammation-induced gene product associated with cachexia [23], while the phenotype of *Adamts1*-null mice also suggests roles for ADAMTS-1 in growth, organogenesis and female fertility [31]. ADAMTS-1 and ADAMTS-8 have both been shown to have anti-angiogenic activity [32], which in the case of ADAMTS-1 may involve binding and sequestration of the angiogenic factor $VEGF_{165}$ by ADAMTS-1 COOH-terminal thrombospondin-like domains [33]. Roles for ADAMTS-like proteinases in fertility and organogenesis in a broad range of species are suggested by the finding that the *gon-1* gene, which encodes an ADAMTS-like product, is essential for gonadal morphogenesis in *Caenorhabditis elegans* [34]. Mutations in the gene for ADAMTS-13 are causal for thrombotic thrombocytopenic purpura, perhaps due to a demonstrated ability to process von Willebrand factor, suggesting ADAMTS-13 to play an important role in human vascular homeostasis [35].

Aside from the role of ADAMTS-2/pNP1 in the biosynthetic processing of procollagens, the phenotype of *Adamts2*-null mice suggests a role in male fertility as well, since male knockout mice were sterile, with an apparent defect in the maturation of spermatogonia [36]. Interestingly, ADAMTS-2 is expressed at dramatically higher levels in seven days post-conception (dpc) murine gas-

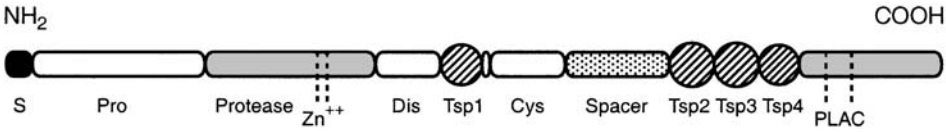

Fig. 2 Shared protein domain structure of ADAMTS-2, 3, and 14. Protein domains are: signal peptide (S); Pro-, Protease, disintegrin-like (Dis), 1st, 2nd, 3rd, and 4th thrombospondin type-1 (Tsp1, 2, 3, and 4), cysteine-rich (Cys), and spacer domains. The *dashed region* within the protease domain represents the Zn^{2+}-binding active site. The *dashed region* within the COOH-terminal domain indicates the relatively conserved PLAC region

trulas than at later gestational times [37], even though mRNAs of the major fibrillar collagens are at relatively low levels at 7 dpc and are found at abundant levels only markedly later in development [38]. This difference in temporal patterns of expression suggests that, in addition to its role as pNPI, ADAMTS-2 is likely to be involved in proteolytic processing of substrates other than the major fibrillar procollagens. Importantly, it has recently been shown that ADAMTS-2 can cleave procollagen type III, in addition to being able to cleave procollagen types I and II [39]. Thus, ADAMTS-2 is both a pNPI and a pNPIII, and the previous distinction made between pNPI and pNPIII as activities of separate proteinases appears to have been a false one. In addition, ADAMTS-3 and -14, which have greater similarities in sequence and domain structure to ADAMTS-2 than do other ADAMTS proteinases, have been shown to have pNPI activity in vitro, and have been suggested as possible sources for the residual pNPI activity observable in the bone, tendon, cartilage, skin, and other tissues of EDS VIIC patients, dermatosparaxic cattle, and *Adamts2*-null mice [37, 40]. It has yet to be determined whether ADAMTS-3 and -14 have both pNPI and pNPIII activities, although in light of the results with ADAMTS-2 [39] and the similar structures of ADAMTS-2, -3, and -14, this seems likely. Thus, ADAMTS-2, -3, and -14 constitute an ADAMTS subfamily that appears to serve as the major, if not sole source of pNP activity in vivo.

2.2 C-Proteinases and the BMP-1/Tolloid Family of Metalloproteinases

If inability to cleave the C-propeptides of fibrillar procollagens is indeed incompatible with fibrillogenesis, then the enzymatic activity responsible for such cleavage could be considered a key control point in morphogenetic processes. The proteinase(s) responsible for this activity might also be considered a reasonable target for therapeutic interventions in the pathological over-deposition of collagenous ECM, such as occurs in the fibroses. Small amounts of procollagen C-proteinase (pCP), the activity that cleaves the C-propeptides of procollagen types I–III, have been purified from the conditioned media of chick embryo tendon organ cultures [11] and cultured mouse fibroblasts [12, 41]. Isolation of sufficient amounts of pCP for NH_2-terminal sequencing of proteolytic fragments showed pCP to be identical to the previously cloned gene product bone morphogenetic protein-1 (BMP-1) [13, 42]. Prior to the identification of BMP-1 as a pCP, peptides derived from BMP-1 had been identified in demineralized bone extracts capable of inducing ectopic endochondral bone formation when implanted into the soft tissues of experimental animals [43]. Extensive purification of this bone morphogenetic activity from demineralized bone extracts found the activity to correspond to fractions containing peptides derived from BMP-1 and from certain members of the transforming growth factor-β (TGF-β) family of proteins, which were designated BMPs-2 through -7 [43–45]. BMP-1 was unlike the TGF-β-related proteins, and had a distinct protein domain structure that included a conserved domain found in the astacin family

of metalloproteinases [46]. In addition, BMP-1 contains an NH_2-terminal prodomain, that must be proteolytically removed for activation to occur [46, 47]; CUB (Complement-Uegf-BMP-1) domains, first characterized in complement components C1r/C1s [48] and thought to mediate protein-protein interactions in various proteins implicated in developmental processes [49]; and an EGF-like motif, that may also be involved in protein-protein interactions [46], but which in some proteins binds Ca^{2+} [50]. Experiments with a series of recombinant truncated forms of BMP-1 have shown that BMP-1 lacking both the EGF motif and the most COOH-terminal CUB domain retains much of its pCP activity [51]. Thus, the functions of the more COOH terminal domains, including additional COOH-terminal domains found on other pCPs (see below), remain to be determined. Such functions may include interactions with substrates other than the major procollagens (see below) and/or targeting to specific areas of the extracellular compartment via binding to specific ECM components.

Although the domain structure of BMP-1 was unique at the time of discovery, BMP-1 has since become the prototype of a family of similarly structured metalloproteinases (Fig. 3) implicated in morphogenetic processes in a broad range of species. Soon after discovery of BMP-1, the *Drosophila* protein Tolloid (TLD) was found to have a domain structure highly similar to that of BMP-1 [52] (Fig. 3). Tolloid, the product of one of at least seven zygotically active genes necessary to formation of the dorsal-ventral axis in early *Drosophila* embryogenesis, acts by potentiating the activity of the TGF-β-like protein Decapentaplegic (DPP) [53, 54]. DPP is more similar in sequence to BMP-2 and BMP-4 than to any other vertebrate TGF-β-like molecule [55]. Moreover, these proteins are functional orthologues, as DPP and BMP-2/-4 are capable of functionally substituting for each other and of subserving the same roles in invertebrate and vertebrate dorsal-ventral patterning [55]. Co-purification of BMP-1 with BMP-2/-4 through a large number of chromatographic steps, suggested molec-

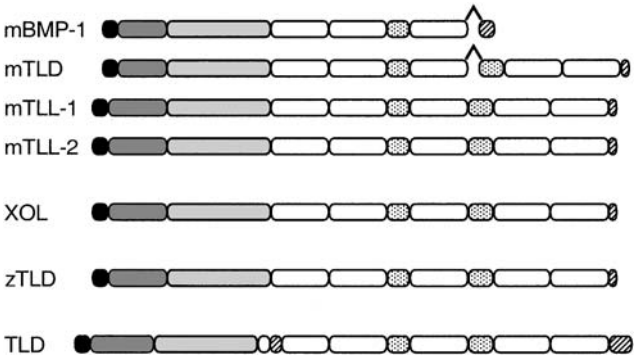

Fig. 3 Domain structures of BMP-1-like proteinases: *Black* denotes signal peptides; *dark gray*, prodomains; *light gray*, protease domains; *white*, CUB domains; *spotted*, EGF-like domains; *hatched*, domains unique to each protein. *Chevrons* denote alternative splicing events that produce BMP-1 and mTLD from the same gene

ular interactions between these molecules, although the nature of such interactions were initially obscure, since proteolytic activation of TGF-β-like proteins is via cleavage at consensus Arg-X-X-Arg sites by furin-like proprotein convertases [56]. Nevertheless, the above observations suggested functional interaction between BMP-1/TLD-like proteinases and DPP/BMP-2/-4 growth factors across a wide range of species. It is now known that TLD acts in dorsal-ventral patterning by cleaving the secreted protein Short gastrulation (SOG) [56], which releases DPP from a latent DPP-SOG complex. Similarly, the BMP-1/TLD-related non-mammalian vertebrate proteinases *Xenopus* Xolloid [57] and zebrafish Tolloid [58] (Fig. 3) are capable of in vitro cleavage of the SOG orthologue *Xenopus* Chordin, and of counteracting the dorsalizing effects of Chordin upon co-overexpression in *Xenopus* [57] or zebrafish [58] embryos.

Mammals have four BMP-1/TLD-related proteinases. The gene that encodes BMP-1, also encodes alternatively spliced mRNA for a second, longer proteinase which, with an additional EGF-like domain and 2 additional COOH-terminal CUB domains, has a domain structure essentially identical to *Drosophila* TLD and which has therefore been designated mammalian Tolloid (mTLD) [59]. In addition, there are two genetically distinct proteinases, designated mammalian Tolloid-like 1 [60] and 2 [61] (mTLL-1 and mTLL-2), with domain structures identical to that of mTLD (Fig. 3). BMP-1 and mTLL-1, but not mTLD or mTLL-2, are capable of cleaving mammalian Chordin in vitro and of counteracting the dorsalizing effects of Chordin upon co-overexpression in *Xenopus* [61]. In addition, although full-length Chordin is not readily detectable in conditioned media of mouse embryo fibroblast (MEF) cultures derived from wild type mice or from knockout mice lacking either the *Bmp1* gene, which encodes both BMP-1 and mTLD, or the Tll1 gene, which encodes mTLL-1, full-length Chordin is readily detectable in media of MEFs derived from embryos doubly homozygous null for both the Bmp1 and Tll1 genes [62]. Together the various data thus demonstrate that BMP-1 and mTLL-1 are together responsible for cleaving Chordin in vivo and that together these two proteinases are therefore involved in potentiating signalling by BMP-2 and -4 in mammalian morphogenetic and homeostatic events.

All four mammalian proteinases have been shown to have pCP activity in vitro [61, 62], with BMP-1 and mTLL-2 demonstrating the most robust and lowest levels of activity, respectively [61, 62]. Moreover, levels of pCP activity are reduced in the culture media of Bmp1-null MEFs [63], whereas pCP activity is undetectable in media of MEFs derived from embryos doubly homozygous null for the Bmp1 and Tll1 genes [62]. Thus BMP-1, mTLD and mTLL-1 are all likely to be physiologically relevant pCPs in mammalian tissues. Absence of detectable pCP activity in Bmp1:Tll1 doubly null MEF media [60] argues against a major role for mTLL-2 as an in vivo pCP, especially since mTLL-2 mRNA is expressed by MEFs [64]. However, small but detectable amounts of apparently mature type I collagen monomers found in ECM associated with Bmp1:Tll1 doubly null MEF cell layers [62] suggest that some residual pCP activity may remain, perhaps provided by mTLL-2. It is also possible that if mTLL-2 does not

play a major role as a pCP in MEF cultures, it may still play significant roles in other cells and tissues.

The above data indicate that at least some of the proteinases that act as pCPs in vivo also serve to modulate the activities of the important morphogens BMP-2 and -4, in vivo. Thus, these proteinases would appear to provide some degree of coordination between formation of the collagenous ECM and BMP signaling in mammalian morphogenetic events.

2.3 Major Procollagen Processing in the Context of Secretion and Fibrillogenesis

Of some interest are the questions of where proteolytic processing of the major procollagens occurs in regard to the collagen producing cell, and when it occurs relative to other steps in the process of fibrillogenesis. BMP-1-like proteinases, ADAMTS-2, and the major fibrillar procollagens are all secreted proteins [4, 39, 65]. Moreover, the neutral pH optima of both pNPs and pCPs [9–12] suggest that these proteinases operate most efficiently in the extracellular space, given the acidic milieu of intracellular compartments. Nevertheless, it has been reported that BMP-1 may, in at least some circumstances, be activated within the trans-Golgi compartment, concomitant with the process of sialylation [47]. It has also been noted that if, as some data suggest, BMP-1-like proteinases are responsible for cleaving the prodomain of the small proteoglycan decorin (see below), then such cleavage way occur within the Golgi apparatus prior to elongation of decorin glycosaminoglycan chains [64]. Thus, although many studies have demonstrated that both pCP activity and unprocessed procollagen are secreted into the extracellular space in cell or organ culture, or demonstrated C-propeptide cleavage in cell-free systems, the possibility exists that some cleavage of procollagen may occur intracellularly in the trans-Golgi compartment.

Classical electron microscopy studies have suggested that fibrillar procollagens form intracellular aggregates, rather than being secreted in monomeric form, and that nascent fibrils may grow within recesses of the fibroblast cell surface via addition of these intracellular aggregates [66, 67]. This view has received support in recent years by studies demonstrating that procollagen forms large electron-dense aggregates in the Golgi, that are too large for secretion via transport vesicles and which are instead transported across the Golgi complex via progressive maturation of the Golgi cisternae [68]. Interestingly, it has been reported that both pCP and pNP activities cleave aggregated procollagen at markedly more rapid rates than they do monomeric procollagen [69]. Thus, it seems likely that major fibrillar procollagens are cleaved as macromolecular aggregates by BMP-1-like proteinases, and it may be that such cleavage occurs either immediately before, or coincident with secretion. The distinction between intracellular and extracellular may be a fine one in this case, since cleavage may occur within fully matured Golgi cisternae as they open into recesses of the fibroblast cell surface, resulting in a pH shift towards neutral in the open cisternae. Certainly, the above speculations are consistent with studies that have

shown the processing of major procollagen C-propeptides to be extremely rapid in tissues, and likely coincident with secretion in a pericellular environment [70, 71]. In contrast to BMP-1-like proteinases, fully activated, mature ADAMTS-2 is not detectable intracellularly, and only appears extracellularly, seemingly bound to extracellular heparan sulfate proteoglycans and perhaps to various components of the ECM [20, 39]. Thus, pNP activities may process procollagen aggregates in a truly extracellular fashion.

2.4 C-Proteinase Enhancer Proteins

Initial characterization of pCP activity from mammalian sources found this activity to be enhanced by a 55-kDa glycoprotein, designated the procollagen C-proteinase enhancer (PCPE), and by 36- and 34-kDa proteolytic fragments of PCPE [41, 72]. PCPE contains two NH_2-terminal CUB domains and a COOH-terminal NTR (netrin) domain [73]. The latter has homology with the COOH-terminal domains of netrins, which are involved in axon guidance; with the NH_2-terminal domains of TIMPs (tissue inhibitors of matrix metalloproteinases); with complement components C3–5; and with the frizzled-related proteins, which are extracellular antagonists of the Wnt family of extracellular ligands [74]. The 36- and 34-kDa PCPE fragments, which contain little or no sequences other than the two CUB domains [73], retain full pCP-enhancing activity and, like the full-length 55-kDa form, bind type I procollagen C-propeptides [41, 72]. Thus, such abilities appear to reside entirely in the CUB motifs. Interestingly, sequence homologies between PCPE and BMP-1 CUB domains were one of the attributes that first suggested BMP-1 as a candidate pCP [73], and these homologies suggest that BMP-1, which like PCPE binds procollagen C-propeptides [13], may do so via its CUB domains. In contrast to activities provided by the PCPE CUB domains, the cleaved COOH-terminal NTR domain, consistent with its homology to TIMPs, has the ability to inhibit MMPs [75]. Thus, PCPE may play a dual role in collagen deposition, by enhancing pCP activity via its CUB domains and by preventing collagen degradation, via MMP inhibition by its COOH-terminal domain. It should be noted, however, that the cleaved PCPE NTR domain inhibits MMP-2 (a gelatinase) with an IC_{50} value of 560 nmol/l, compared to an IC_{50} value of 1.6 nmol/l for TIMP-2, suggesting that a metalloproteinase(s) other than the MMPs, or at least other than MMP-2, may be the major target of this fragment [75]. It is interesting to speculate that whatever functions are provided by full-length PCPE, it may serve as a precursor from which functional NH_2- and COOH-terminal products are derived. PCPE may also serve additional roles, since disruption of the rat PCPE gene can result in anchorage-independent growth and loss of contact inhibition in cultured fibroblasts [76], although a mechanism for these latter observations is not clear.

PCPE has no intrinsic pCP activity but, rather, appears to act by binding procollagen with a 1:1 stoichiometry and, in so doing, facilitating procollagen C-propeptide cleavage, perhaps via inducing conformational changes in the

substrate [72, 77]. The enhancing activity of PCPE seems specific to pCP activity and does not extend to pNP activity, or to cleavage of other substrates of BMP-1/Tolloid-related proteinases, such as Chordin or type V procollagen (BM Steiglitz and DS Greenspan, unpublished data). Recently, a second PCPE (PCPE2) has been described, with a protein domain structure and levels of pCP-enhancing activity similar to those of PCPE (redesignated PCPE1) [78, 79]. The two PCPEs differ widely in their distributions of expression: PCPE1 has a broad distribution of expression throughout developing mesenchymal tissues, with especially high levels in ossifying bone [73, 79], while developmental expression of PCPE2 seems primarily localized to non-ossified cartilage [79]. Nevertheless, although PCPE1 is predominantly expressed in tissues rich in type I collagen and PCPE2 is predominantly expressed in cartilage, in which type II collagen is the major fibrillar collagen, the two PCPEs are equally efficient in enhancing cleavage of procollagen I and II C-propeptides [79]. Surprisingly, both PCPE1 and 2 were found capable of binding collagen fibrils in tissues and to collagen molecules devoid of C-propeptides in vitro [79]. The same studies also showed pCPs such as mTLL-1 to be capable of binding the triple helical portions of collagen monomers at the same types of sites used by PCPE1 and 2. Previous studies had shown that PCPE1 increases the maximal velocity (V_{max}) of the pCP reaction, but that it also decreases the apparent K_m [72]. The observation that, at the relatively high ratios of PCPEs and pCPs to procollagen used in in vitro assays, PCPEs and pCPs compete for similar sites on the collagen triple helix [72] can explain how PCPEs lower the apparent K_m of the pCP reaction in such assays. K_D values for binding of PCPE1 to procollagen I and III C-propeptides have been estimated at ~170 and ~370 nmol/l, respectively [77], while K_D values for PCPE binding to the collagen triple helix [68] and to type I procollagen [77] have been estimated at ~10 and ~1 nmol/l, respectively. Thus, although PCPE1 was first isolated, in part, via its ability to bind the type I procollagen C-propeptide [72], the majority of procollagen-binding activity may derive from binding of PCPEs to the collagen triple helical domain. The apparent binding of both PCPEs and pCPs to a number of low affinity binding sites on the collagen triple helix may act to increase local concentrations of the factors required for biosynthetic processing by limiting their diffusion away from procollagen substrate, thus facilitating pCP activity at the C-propeptide cleavage site.

2.5 Noncollagenous Substrates of the BMP-1/Tolloid-Related C-Proteinases

The sites at which BMP-1-related proteinases cleave procollagens I–III and Chordin have certain similarities. Most notably, each site contains an Asp residue at position P1′ and a residue with an aromatic side chain (i.e. Tyr or Phe) and/or Met residues NH_2-terminal to the cleavage site, usually in position P3 or P2 (Fig. 4). Numerous extracellular proteins are synthesized as precursors with prodomains that are proteolytically cleaved during biosynthesis, to yield the mature functional protein. A subset of these precursors have in vivo cleavage sites

Procollagen C-propeptides	
α1(I)	DGGR**YY**RA...**D**DANVVRD
α2(I)	**Y**DGD**FY**RA...**D**QPRSAPS
α1(II)	DPLQ**Y**MRA...**D**QAAGGLR
α1(III)	GG**F**AP**YY**G...**D**EPMDFKI
α2(V)	DPLPE**F**TE...**D**QAAPDDK
α1(VII)	RPLPS**Y**AA...**D**TAGSQLH
Laminin 5 γ2 chain	DTGDC**Y**SG...**D**ENPDIEC
Lysyl oxidase	SHVDR**M**VG...**D**DPYNPYK
SLRP's	
Biglycan	DDGP**FMM**N...**D**EEASGAD
Decorin	GL**F**D**FM**LE...**D**EASGIGP
Osteoglycin	EKSLQLQK...**D**EVIPSLP
Epiphycan	TEL**F**N**Y**DS...EVYDAILE
Chordin	
N-terminal site	DPEHRS**Y**S...**D**RGEPGVG
C-terminal site	KLGDP**M**QA...**D**GPRGCRF
SIBLING's	
DMP1	**F**DDEG**M**QS...**D**DPESTRS
DSPP	
Myostatin	IDQ**Y**DVQR...**D**DSSDGSL
Procollagen N-propeptides	
α1(V)	AVPDTP**Q**S...**Q**D**P**NPDEY
α1(XI)	SAPKAA**Q**A...**Q**E**P**QIDEY
α2(XI)	GQRERP**Q**N...**Q**Q**P**HRAQR

Fig. 4 Known and potential cleavage sites of BMP-1-like proteinases. Residues in *boldface* are conserved at the cleavage sites of either procollagen N-propeptides, or other substrates

that resemble the cleavage sites at which mammalian BMP-1-like proteinases cleave procollagens I–III and Chordin. Such precursors have become candidates, and a number have been demonstrated to be, substrates of BMP-1-related proteinases. Known noncollagenous substrates are presented below.

2.5.1 Small Leucine-Rich Proteoglycans

Biglycan and decorin are small nonaggregating proteoglycans that contain chondroitin sulfate or dermatan sulfate side chains and belong to the family of small leucine-rich proteoglyans (SLRPs) of the ECM [80]. The phenotypes of mice homozygous null for biglycan or decorin have shown the former to be a positive regulator of bone growth [81] and the latter a regulator of type I collagen fibrillogenesis in skin and tendon [82]. High levels of expression in preosteogenic cells and a pericellular distribution are consistent with a role for biglycan in osteoblast differentiation, whereas association of decorin expres-

sion with tissues rich in fibrillar collagens is consistent with a role in fibrillogenesis [83]. Although the molecular bases for the biological roles of biglycan and decorin are unclear, they may involve demonstrated abilities of the two to interact with various collagens, other ECM proteins, and TGF-β [84–90]. Biglycan and decorin are synthesized as pro-forms containing NH_2-propeptides of 21 and 14 residues, respectively, that are completely removed in most, but not all connective tissues [91]. Residues at the human probiglycan and prodecorin in vivo cleavage sites, M(M/L)N-DEE and M(L/I)E-DE(A/G), are conserved in various species [91–96] and show similarities to the procollagen I–III C-propeptide cleavage sites (Fig. 4), which made these proteins candidate substrates for the mammalian BMP-1-related proteinases. Studies have since shown that the pro-form of biglycan is efficiently cleaved in vitro by BMP-1 at the physiologically relevant maturation site and not at any other site [64]. Related proteinases mTLD and mTLL-1 have lower levels of the same activity, whereas mTLL-2 lacks detectable levels of such activity [64]. Moreover, although only mature biglycan, is detectable in wild type MEF conditioned media, a mixture of ~40% mature biglycan and 60% probiglycan is detectable in the media of Bmp1-/-MEFs, and only uncleaved probiglycan is detectable in culture media of MEFs doubly null for both the Bmp1 and Tll1 genes [64]. Thus, products of the Bmp1 and Tll1 genes are together responsible for all detectable probiglycan-processing activity in vivo, or at least in MEFs. Clearly, biosynthetic processing of biglycan, a positive regulator of bone growth, conforms with the roles of BMP-1, mTLD, and mTLL-1 as pCPs, an activity necessary to formation of the collagenous ECM of bone; and as activators of signaling by TGF-β-like BMPs, first characterized as inducers of bone formation. Such observations suggest BMP-1-like proteinases to be central in orchestrating events involved in bone formation.

Although biosynthetic processing of decorin, a modulator of type I collagen fibrillogenesis, would also conform to the known roles of the BMP-1-related proteinases, it has yet to be determined which proteinase(s) is responsible for decorin processing. It must be noted, however that if prodecorin and probiglycan are processed by the same proteinases, then it is with different kinetics, since transfected 293-EBNA cells that secrete only unprocessed recombinant probiglycan, secrete only fully processed mature recombinant decorin [64].

The SLRPs can be divided into four classes, based on sequence homologies and protein domain structures [80, 97]. Class I comprises decorin and biglycan, which show higher sequence homology to each other than is found in any other pair of SLRPs; class II comprises fibromodulin, lumican, keratocan, PRELP, and osteoadherin; class III consists of chondroadherin; while class IV comprises epiphycan and osteoglycin. While there is no evidence that class II or III SLRPs are processed from precursors to mature forms, in vivo biosynthetic processing has been reported for the two class IV SLRPs, osteoglycin [98] and epiphycan [99]. Very recent data show that osteoglycin has a prodomain that is cleaved, with varying efficiencies, by all four mammalian BMP-1-related proteinases in vitro and that this prodomain is processed by wild type MEFs, but not by MEFs doubly null for the Bmp1 and Tll1 genes [100]. Thus, at least three of the mammalian

BMP-1-related proteinases are involved in the in vivo biosynthetic processing of osteoglycin. This activity appears to conform with other activities of the BMP-1-like proteinases, since osteoglycin appears to play a role in regulating collagen fibrillogenesis [101]. Future studies are needed to determine whether precursors of other SLRPs, such as epiphycan are processed by BMP-1-related proteinases.

2.5.2 Lysyl Oxidase

Lysyl oxidase is a secreted amine oxidase necessary for formation of the covalent cross links that give collagen and elastic fibers much of their tensile strength. It does this by oxidatively deaminating ε-amino groups of certain peptidyl lysine and hydroxylysine residues within collagen monomers and peptidyl lysines in elastin precursors. This results in conversion of the lysines to peptidyl-α-aminoadipic-δ-semialdehydes and the hydroxylysines to peptidyl-δ-hydroxy-α-aminoadipic-δ-semialdehydes, with subsequent spontaneous condensation of these peptidyl aldehydes with other vicinal peptidyl aldehydes or with unreacted ε-amino groups to form intra- and intermolecular covalent cross-links [102]. Lysyl oxidase is synthesized as a zymogen or proenzyme, with a prodomain that must be cleaved to produce the mature active form [103]. Because the site for cleavage contains a P1′ Asp (Fig. 4), pCP activity from chick embryo tendon organ culture was examined for its ability to activate lysyl oxidase [104]. It was found to cleave the proenzyme to produce mature active lysyl oxidase with an NH_2-terminus identical to that of mature lysyl oxidase cleaved in media of arterial smooth muscle cultures [104]. Subsequently it has been shown that all four mammalian BMP-1-related proteinases are capable of cleaving prolysyl oxidase in vitro, and solely at the physiologically appropriate site [105]. Moreover, in contrast to wild type MEFs, MEFs doubly null for the Bmp1 and Tll1 genes produce predominantly unprocessed prolysyl oxidase and their media has lysyl oxidase activity only 30% that of wild type [105]. Thus, Bmp1 and Tll1 gene products are responsible for the majority of proteolytic activation of lysyl oxidase, at least in MEF cultures, illustrating another way in which the mammalian BMP-1-related proteinases affect ECM formation.

2.5.3 SIBLING Proteins

Noncollagenous proteins of the SIBLING (Small Integrin-Binding LIgand N-linked Glycoprotein) family, which includes dentin sialophosphoprotein (DSPP) and dentin matrix protein-1 (DMP1), are secreted and deposited into bone and dentin ECM during assembly and mineralization, and are thought to initiate mineralization through their acidic calcium binding domains [106–110]. Cleaved COOH-terminal domains of DSPP and DMP1 are found in extracts of mineralized tissues [111, 112], and are capable of stimulating in vitro hydroxyapatite crystal formation [113, 114], prompting suggestions that in vivo

processing of full-length DSPP and DMP1 occurs, yielding bioactive cleavage products. The NH_2-terminus of the cleaved DSPP COOH-terminal domain extracted from dentin, shows in vivo cleavage to have occurred between residues Gly^{447} and Asp^{448} [111], numbering from the initial Met of full-length preproD-SPP. Analysis of DMP1-derived cleavage products extracted from bone shows in vivo cleavage of DMP1 to have occurred at multiple sites, only one of which, between Ser^{196} and Asp^{197}, occurs within sequences strictly conserved within DMP1 in a broad range of species [112]. Similarities between the described cleavage sites of DSPP and DMP1 and previously characterized substrates (Fig. 4), suggested these two proteins as candidate substrates for BMP-1-related proteinases. Experiments have shown recombinant human DMP1 to be cleaved to varying extents by all four mammalian BMP-1-related proteinases, yielding fragments similar in size to those isolated from bone [115]. Moreover, this cleavage occurs between Ser^{196} and Asp^{197}, the most conserved site within DMP1 sequences from a wide range of species [115]. In addition, processing of DMP1 is deficient in Bmp1-/-;Tll1-/- doubly homozygous null MEF cultures [115], further supporting the possibility of roles for BMP-1-related proteinases in the in vivo processing of DMP1. It is yet to be determined whether DSPP is similarly processed by BMP-1-related proteinases. Nevertheless, it is conceptually attractive that the same proteinases that regulate biosynthesis of type I collagen, the major structural protein of bone, and that activate TGF-β-like BMPs responsible for osteoblastic differentiation, may also regulate activities of SIBLING proteins important to the mineralization of bone and other hard tissues.

2.5.4 Laminin-5

Laminin-5, a non-collagenous heterotrimer composed of α3, β3, and γ2 chains, is a major component of anchoring filaments within the basement membrane at the dermal-epidermal junction and is important in stabilizing attachment of epithelial hemidesmosomes to basement membranes [116]. Null mutations in any of the three genes encoding the laminin-5 chains lead to the lethal blistering genetic disease Herlitz's junctional epidermolysis bullosa [117].

Cultured keratinocytes secrete a laminin-5 precursor with 200 kDa α3 and 155 kDa γ2 chains, that are processed upon secretion to yield 165 and 105 kDa forms, respectively, corresponding to the sizes of α3 and γ2 chains isolated from tissues [116, 118]. Exogenously added MMP-2 is capable of cleaving the γ2 chain of rat laminin-5 [119], albeit at a site not conserved in the human laminin-5 γ2 chain [120]. Additional data have suggested that membrane type 1 matrix metalloprotease (MT1-MMP) may be the proteinase primarily responsible for cleavage of the rat laminin-5 γ2 chain, and that MMP-2 may play an ancillary role [121]. However, only BMP-1-related proteinases have been shown to cleave the human γ2 chain at the authentic site at which it is cleaved by keratinocytes [122]. BMP-1-related proteinases, MMP-2, MT1-MMP, and other proteinases are capable of cleaving the α3 chain to produce a form indistinguishable in size

from the keratinocyte processed form [120, 122, 123]. However, a small molecule inhibitor, that seems highly specific for the BMP-1-related proteinases, inhibits cleavage of α3, as well as γ2 chains in keratinoycyte cultures. In addition, neither MT1-MMP nor MMP-2 cleave the human laminin-5 γ2 chain, whereas processing of laminin-5 appears to be deficient in the skin of Bmp1 null mice [120]. All four BMP-1-related proteinases are capable of cleaving both the α3 and γ2 chains [120], although two studies have suggested mTLD as the major BMP-1-related proteinase secreted by keratinocytes [65, 120]. Thus, considerable evidence supports roles for BMP-1-related proteinases, and perhaps mTLD, in processing laminin-5 in vivo, although, the question of which proteinase(s) cleave laminin-5 in vivo should still be considered controversial at this time. The question is of considerable importance, as the proteolytically processed form of laminin-5 is more conducive to epithelial cell migration than is the unprocessed form [119], suggesting that targeting the proteinase(s) responsible for such processing may be an effective therapeutic approach towards limiting invasiveness of carcinomas.

2.5.5 Myostatin/GDF-8

Myostatin, also known as growth differentiation factor 8 (GDF-8), is a TGF-β-like protein that is expressed almost exclusively in cells of myogenic lineage throughout development and in skeletal muscle in the adult, and it negatively regulates skeletal muscle mass [124–127]. Although the myostatin prodomain, like the prodomains of all TGF-β-like molecules, is cleaved by furin-like proprotein convertases to yield the mature active molecule, the cleaved myostatin prodomain remains non-covalently associated with mature myostatin as a latent complex [124–126, 128]. Recently it has been shown that BMP-1, mTLL-1, and mTLL-2 cleave within myostatin prodomain sequences, to liberate active mature myostatin from the latent complex [129]. Lower levels of activity were shown for mTLD [129]. The cleavage site resembles those of a number of other substrates of BMP-1-like proteinases, including existence of a P1′ Asp [129] (Fig. 4). Substitution of the myostatin P1′ Asp with Ala, makes the prodomain impervious to cleavage by BMP-1-like proteinases and prevents activation of myostatin latent complexes in vitro. Importantly, systemic administration of the mutant myostatin prodomain to adult mice induces marked increases in muscle mass, whereas wild type prodomain does not have this effect [129]. Thus, cleavage at this site is an important activator of myostatin activity in vivo. While it is not clear which BMP-1-like proteinase may activate myostatin in vivo, it is of interest that mTLL-2 is specifically expressed by skeletal muscle during embryonic development, whereas BMP-1 has a broad developmental distribution of expression that includes muscle [61]. It is intriguing to speculate whether such proteinases may operate in disease processes such as muscular dystrophies, via potentiating fibrotic tissue infiltration while negatively regulating skeletal muscle regeneration.

3
Processing of the Minor Fibrillar Collagens

The minor fibrillar collagens comprise low abundance fibrillar collagen types V and XI. Monomers of collagens V and XI are incorporated into growing fibrils of the much more abundant major fibrillar collagens I and II, respectively, and regulate the sizes and shapes of the resultant heterotypic fibrils [130–136]. Type V collagen is most widely distributed in tissues as a heterotrimer of the chain composition $\alpha1(V)_2\alpha2(V)$ [137], but is found in a limited number of cell types and tissues as a rare $\alpha1(V)_3$ homotrimer [138–140]. In addition, a poorly characterized $\alpha1(V)\alpha2(V)\alpha3(V)$ heterotrimer has been isolated primarily from human placenta [141, 142], but has also been reported in uterus, skin, and synovial membranes [137, 143–145]. Localization of $\alpha3(V)$ RNA within nascent ligamentous attachments of developing joints, membranous linings of developing skeletal muscle, and developing and regenerating peripheral nerves in mouse and rat [146, 147], suggests roles for $\alpha1(V)\alpha2(V)\alpha3(V)$ heterotrimer in these tissues as well. Type XI collagen, in the form of an $\alpha1(XI)\alpha2(XI)\alpha3(XI)$ heterotrimer [148], was first characterized as a minor collagen of cartilage. However, findings of type XI chains in noncartilaginous tissues [150], of type V chains in cartilage [150], and of cross-type heterotrimers composed of both type V and XI chains [151, 152] suggest that type V and XI chains constitute a single collagen type in which different combinations of chains associate in a tissue-specific manner. Unlike the major fibrillar collagens I–III, processed collagens V and XI retain residual N-propeptide sequences [139, 140, 153–157]. These appear to be of functional importance since, as shown for collagen V, they protrude beyond the surface of heterotypic fibrils and may control fibrillogenesis by sterically hindering further addition of collagen monomers to the fibril surface [155]. The minor fibrillar procollagens are similar in domain structure to the major fibrillar procollagens, in consisting of a central collagenous domain of ~1000 amino acid residues in length, bracketed by N- and C-propeptides [137]. However, the pro-$\alpha1(V)$, pro-$\alpha1(XI)$, pro-$\alpha2(XI)$, and pro-$\alpha3(V)$ chains, more similar in sequence and domain structure to each other than to other fibrillar procollagen chains, differ most from the other procollagen chains in the configurations and large sizes of their N-propeptides (Fig. 5) [158–164]. The genes for these procollagen chains also share a conserved intron-exon structure that is diverged from the conserved intron-exon structure of the major fibrillar procollagen chain genes [161, 162, 165, 166], implying at least two evolutionary pathways that have led to distinct subclasses of fibrillar collagen chains. In contrast, domain structure of the pro-$\alpha2(V)$ chain (Fig. 5) and the structure of its cognate gene resemble those of the major fibrillar collagens [167, 168].

Low levels of minor fibrillar procollagens in tissues and cell cultures have limited their characterization. Thus, to facilitate characterization of type V procollagen processing, recombinant pro-$\alpha1(V)_3$ homotrimers were produced for in vitro cleavage assays [169]. Recombinant pro-$\alpha1(V)_3$ homotrimers were used for initial studies, as their production was more straightforward than produc-

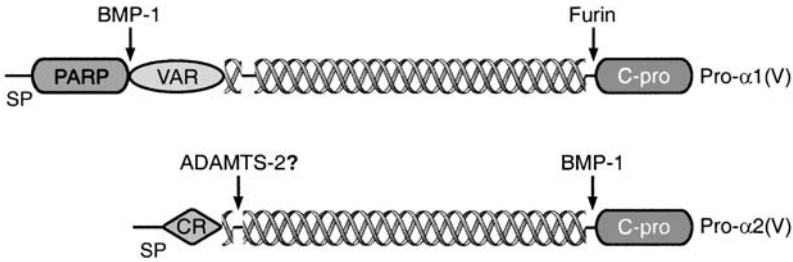

Fig. 5 Processing of type V procollagen. SP denotes signal peptides. PARP and variable (VAR) subdomains of the pro-α1(V) NH_2-terminal globular sequences are shown. C-pro denotes C-propeptides, and CR denotes the cysteine-rich subdomain of the pro-α2(V) N-propeptide

tion of recombinant pro-$\alpha1(V)_2$pro-α2(V) heterotrimers, although the latter are more representative of the majority of type V procollagen in vivo. Surprisingly, secretion of recombinant pro-$\alpha1(V)_3$ homotrimers from cultures of transfected human 293-EBNA cells was accompanied by cleavage of pro-$\alpha1(V)_3$ C-propeptides at a consensus (R/K)XRR site suitable for cleavage by furin-like proprotein convertases [169]. This cleavage was partially inhibited by culturing cells in the presence of 100 mmol/l L-arginine, resulting in intact pro-$\alpha1(V)_3$ homotrimers that could indeed be cleaved in vitro by a soluble form of furin, and solely at the site used in cell cultures [169]. Incubation of intact pro-$\alpha1(V)_3$ homotrimers with BMP-1 did not lead to C-propeptide cleavage, but instead led to cleavage at a single site within pro-α1(V) NH_2-terminal globular sequences [169] (Fig. 5). Non-cleavage of the C-propeptide by BMP-1 was surprising, since the domain structure of pro-α1(V) C-propeptides is similar to that of the major fibrillar procollagen C-propeptides, and because the pro-α1(V) COOH-telopeptide, the short linker region that connects the main collagenous domain to the C-propetide, contains three Asp residues [158, 159], one of which might have marked the P1′ residue of a pCP cleavage site. Also surprising was the finding that the BMP-1 cleavage site within pro-α1(V) NH_2-terminal globular sequences differed from previously characterized cleavage sites of BMP-1-related proteinases in possessing Gln rather than Asp in the P1′ position, and that other residues flanking the scissile bond lacked resemblance to residues flanking scissile bonds in previously identified substrates of BMP-1-like proteinases [169] (Fig. 4). Nevertheless, the α1(V) NH_2-terminus produced by cleavage at this site corresponds to an equivalent NH_2-terminus on the processed ECM form of the similar α1(XI) collagen chain [156], consistent with physiological significance for cleavage at this site. Conservation of residues surrounding the NH_2-terminal pro-α1(V) BMP-1 cleavage site (Fig. 4) and of furin proprotein convertase recognition sites in COOH-telopeptide domains in the pro-α1(XI) and pro-α2(XI) chains, suggests that these related chains may be processed in a fashion similar to that of the pro-α1(V) chain.

A subsequent study with recombinant pro-$\alpha1(V)_3$ homotrimers showed that blocking cleavage of the C-propeptide in cell cultures, by use of the highly spe-

cific furin inhibitor decanoyl-RVKR-chloromethyl ketone, results in full-length pro-α1(V)$_3$ homotrimers that can be cleaved in vitro by BMP-1 at both the NH_2-terminal globular site identified in the previous study, and within the COOH-telopeptide domain at a site with a P1′ Asp [170]. This second pro-α1(V)$_3$ homotrimer study, which used higher levels of BMP-1 activity than the first, nevertheless found cleavage at the COOH-telopeptide site to be markedly less efficient than cleavage at the site in NH_2-terminal globular sequences. In a third study, recombinant pro-α1(V)$_2$pro-α2(V) heterotrimers were successfully produced and employed in demonstrating that pro-α1(V) NH_2-terminal globular sequences are cleaved by BMP-1 at the same site as in pro-α1(V)$_3$ homotrimers, and that pro-α1(V) C-propeptides are still cleaved by furin in the context of pro-α1(V)$_2$pro-α2(V) heterotrimers [171]. In contrast, the C-propeptides of pro-α2(V) chains within the same pro-α1(V)$_2$pro-α2(V) heterotrimers were not cleaved by furin, but were instead cleaved by BMP-1 at a site similar to those at which the C-propeptides of the major fibrillar procollagens are cleaved, in that it contained a P1′ Asp and a P3 Phe (Figs. 4 and 5). Thus, furin cleavage of procollagen C-propeptide sequences seems confined to procollagen chains of the pro-α1(V)-like subclass, a likelihood supported by the lack of furin cleavage consensus sequences in all of the major fibrillar procollagen COOH-telopeptide regions, and the demonstration that neither procollagen I or II are susceptible to cleavage by furin in vitro [171].

In vivo significance of the patterns observed for in vitro processing of recombinant type V procollagens by furin- and BMP-1-like proteinases was ascertained by comparing processing of endogenous pro-α1(V) chains in cultures of wild type and Bmp1;Tll1 doubly homozygous null MEFs [171]. In such experiments, media of wild type cultures were found to contain pN-α1(V) chains (from which C-propeptides had been removed, but which still retained full, uncleaved NH_2-terminal globular sequences) and mature α1(V) chains (corresponding to the matrix form in which both C-propeptide and PARP subdomain of the NH_2-terminal globular domain have been removed) (Fig. 5), while media of Bmp1;Tll1 doubly null MEFs contained only the pN-α1(V) form [171]. Such results are wholly consistent with the probability that products of the Bmp1 and Tll1 genes are responsible for cleaving pro-α1(V) N-propeptides, but not C-propeptides in vivo. In addition, culturing of MEFs in the presence of the specific furin inhibitor decanoyl-RVKR-chloromethyl ketone resulted in detection of only intact full-length pro-α1(V) chains and pC-α1(V) forms (which retain the C-propeptide, but from which the PARP subdomain of the NH_2-terminal has been removed) (Fig. 5) in wild type media, and only intact full-length pro-α1(V) chains in the media of Bmp1;Tll1 doubly null MEFs [171]. These latter results are consistent with the probability that, whereas products of the Bmp1 and Tll1 genes are responsible for cleaving pro-α1(V) NH_2-terminal sequences in vivo, furin-like activity is solely responsible for cleaving pro-α1(V) C-propeptides in vivo.

The relative lack of similarity between residues flanking the BMP-1 cleavage site in pro-α1(V) NH_2-terminal globular sequences and previous sites cleaved

by BMP-1-like proteinases indicates that BMP-1-like proteinases, like other astacin-like proteinases [46], are not highly specific for such residues and suggests a reappraisal of features that influence cleavage of substrates by BMP-1-like proteinases. In fact, it should be stressed that even the limited conservation found in residues flanking cleavage sites in previously characterized substrates of BMP-1-like proteinases (Fig. 4) is somewhat misleading, since 1) conserved residues flanking sites for cleavage of C-propeptides of procollagen I-III, and pro-α2(V) chains sites may reflect the similar evolutionary origin of these chains rather than a requirement for such residues for cleavage, and 2) a number of other substrates have been selected as candidate substrates based on the similarities of their in vivo cleavage sites to those of procollagens I–III. Nevertheless, it should be noted that in at least some cleavage sites a P1′ Asp is essential for cleavage, since substitution of the P1′ Asp by Ala at either the pro-α2(I) C-propeptide cleavage site [172] or the site within myostatin prodomain sequences [129] makes these sites impervious to cleavage by BMP-1-like proteinases.

Cleavage products of known substrates of BMP-1-like proteinases, such as procollagen I [62], Chordin [62], and probiglycan [64], are absent from media of Bmp1;Tll1 doubly null MEFs. This suggested that cleavage products of other substrates would be absent as well. If so, such fragments, represented as spots on 2D gels of wild type MEF media, would be absent from 2D gels of Bmp1;Tll1 null MEF media. Thus, identification of such spots via a proteomics-based approach could lead to identification of novel substrates in a fashion that, unlike the previous candidate approach, would be unbiased by preconceived notions of features that a cleavage site for BMP-1-like proteinases must have [62]. In such a proteomics approach, four spots on a 2D gel of wild type MEF media, but missing from a 2D gel of Bmp1;Tll1 null MEF media, were subjected to analysis by mass spectrometry [62]. Three of the spots represented the C-propeptides of the pro-α1 and pro-α2 chains of procollagen I and of the pro-α1 chain of procollagen III, thus validating the approach as a means for identifying in vivo substrates of Bmp1 and Tll1 gene products. The fourth spot represented the PARP subdomain of pro-α1(XI) NH_2-terminal globular sequences [62]. Moreover, subsequent immunoblot analyses of materials from MEF media supported the conclusion that the pro-α1(XI) NH_2-terminal globular domain is indeed processed by BMP-1-like proteinases and that Furin-like proteinases are solely responsible for cleaving pro-α1(XI) C-propeptides in vivo [62].

An independent study, using recombinant proteins for in vitro assays, also demonstrated that BMP-1 cleaves within pro-α1(XI) NH_2-terminal globular sequences [173]. The latter study also showed that pro-α1(XI) isoforms, differing in NH_2-terminal globular variable subdomain (Fig. 5) sequences due to alternative splicing [174, 175], differ in the efficiency with which they are processed by BMP-1, although each isoform is cleaved at the same site [173]. It was speculated that such differences in processing rates may result in fine regulation of interactions between heterotypic type II/XI collagen fibers and other ECM components, as such interactions may be mediated by α1(XI) NH_2-termini [173]. Peptides extracted from fetal calf cartilage suggest additional proteolytic

processing of pro-α1(XI) NH_2-terminal sequences by proteinases other than BMP-1-like proteinases, at various sites [173], although it is not clear whether these additional cleavages are physiologically relevant or represent artifactual cleavages that occurred during tissue extraction. It should be noted that alternative splicing of variable subdomain sequences has been observed for pro-α1(XI) and pro-α2(XI) chains [162, 174, 175], but does not appear to occur in pro-α1(V) chains [158].

Conservation of sequences for cleavage by furin-like proproteins within the pro-α3(V) COOH-telopeptide [146], suggests that cleavage of this domain will be by furin-like enzymes. However, although residues at the P1′, P3′, and P2 positions of the pro-α1(V) NH_2-terminal cleavage site are conserved at corresponding positions in pro-α1(XI) and pro-α2(XI) NH_2-terminal globular sequences [169], such conservation is lacking in pro-α3(V) [146]. Thus, it remains to be determined whether the pro-α3(V) NH_2-terminal globular domain is processed and, if so, at which location and by which proteinase(s). Another unresolved aspect of the proteolytical processing of the minor fibrillar procollagens concerns processing of the pro-α2(V) N-propeptide. It has been suggested that the pro-α2(V) N-propeptide may be cleaved by the same pNP activity that cleaves N-propeptides of major fibrillar procollagens, primarily because of similarities in the domain structure of the pro-α2(V) and procollagen I–III N-propeptides, but also because pro-α2(V) NH_2-terminal globular sequences contain the sequence Phe-Ser-Ala-Gln at a position similar to the Phe/Tyr-Ser/Ala/Pro↓Gln sites at which pNP activity cleaves the N-propeptides of procollagens I-III [168]. However, examination of the mature matrix form of α2(V) chains from various tissues has suggested that NH_2-globular sequences are retained in their entirety [140]. Thus, it will be of interest to determine whether the pro-α2(V) N-propeptide is susceptible or resistant to cleavage by ADAMTS-2.

Cleavage of the C-propeptides of pro-α1(V)-like chains by furin-like proteinases may add another level of regulation to collagen fibrillogenesis and link fibrillogenesis to the many other processes governed by these enzymes, which represent the major processing enzymes of the constitutive secretory pathway [176, 177]. In its native form, furin has a transmembrane domain and is primarily localized to the trans-Golgi, but cycles between the tran-Golgi and plasma membrane [176, 177]. Thus, cleavage of pro-α1(V) C-propeptides may occur in the trans-Golgi or at the cell surface. This would agree with early observations that cleavage of pro-α1(V) C-propeptides is relatively rapid [137]. However, furin may also be released from the plasma membrane as a secreted form [178], and thus extracellular cleavage of pro-α1(V) C-propeptides by furin-like activities cannot be excluded.

Cleavage of the C-propeptides of pro-α2(V) chains, which by various criteria are more similar to procollagen I-III chains than to pro-α1(V) chains, by BMP-1-like proteinases, and the likely cleavage by BMP-1-like proteinases of the C-propeptide of the pro-α3(XI) chain, a modified product of the type II collagen pro-α1(II) chain gene [150], may serve to coordinate deposition of minor and major fibrillar collagen monomers within heterotypic fibrils.

Quite recently, sequences for two novel fibrillar-like procollagen chains, proα1(XXIV) and pro-α1(XXVII), have been mined from the genome databases [179, 180]. Each of these chains shows certain similarities to the pro-α1(V), pro-α1(XI), pro-α2(XI) and pro-α3(V) chains, and it will be of interest to determine at which sites and by which proteinases they are processed.

4 Processing of Type VII Collagen

Type VII collagen, a nonfibrillar collagen composed of three identical α1(VII) chains, is the major and perhaps sole component of anchoring fibrils, which are involved in attachment of external epithelia, such as epidermis, to underlying stroma [181] (Fig. 6). Type VII collagen is produced as precursor procollagen molecules [182]. Upon secretion, these associate into antiparallel dimers, C-propeptides are cleaved, and anchoring fibrils are believed to form via lateral association of type VII collagen dimers into macromolecular structures [182–184]. The functional importance of anchoring fibrils is evident in the severe blistering phenotype of dystrophic epidermolysis bullosa (DEB) that results from defective or absent anchoring fibrils, resulting from mutations in the type VII collagen gene [185–187]. The procollagen C-propeptide, also known as the NC-2 domain, appears necessary for initial formation of type VII collagen antiparallel dimers and also contains Cys residues necessary to formation of disulfide bonds that stabilize the dimer [188, 189].

Proteolytic removal of procollagen VII C-propeptides seems necessary for proper formation of anchoring fibrils, as an in-frame exon-skipping mutation that removes 29 amino acid residues containing the in vivo cleavage site results in DEB [190]. Since BMP-1-like proteinases appear to be involved in laminin 5

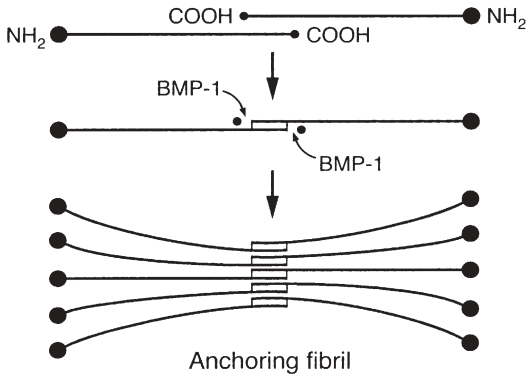

Fig. 6 Processing of type VII procollagen C-propeptides, and lateral assembly of antiparallel dimers to form anchoring fibrils. *Vertical lines* connecting monomers represent disulfide bonding

processing [120, 122], which like collagen VII is located at the dermal-epidermal junction, and since the 29 residues containing the procollagen VII C-propeptide cleavage site contain sequences similar to those at a majority of sites cleaved by BMP-1-like proteinases (Fig. 4), procollagen VII became a candidate substrate for BMP-1-like proteinases. In vitro assays have shown that BMP-1 and mTLL-1 are capable of cleaving a truncated recombinant form of procollagen VII at the predicted site (Fig. 4), with lesser levels of this activity noted for mTLL-2 (mTLD was not tested) [191]. Consistent with an in vivo role for such activity, BMP-1, which is capable of cleaving authentic procollagen VII from normal keratinocytes, is incapable of cleaving mutant procollagen VII from which 29 residues containing the cleavage site have been deleted, and which remains uncleaved in tissues of DEB patients. However, collagen VII in the skin of Bmp1-null 17.5 dpc fetuses appears to be processed and deposited in skin to the same extent as in wild type littermates. It remains unclear at this time whether this observation argues against a role for in vivo processing of procollagen VII by products of the Bmp1 gene, or whether functional redundancy by mTLL-1 provides sufficient residual activity to explain the processing of procollagen VII in Bmp1-null skin. The latter possibility cannot be directly tested at this time, as Tll1 null and Bmp1;Tll1 doubly null embryos die on or prior to 14 dpc [62, 192], at which time the skin basement membrane zone is undeveloped and unamenable to immunohistological or biochemical analysis. It is also possible that mTLL-2 or an unrelated proteinase(s) contributes to in vivo processing of procollagen VII. Creation of conditional knockout mice, in which effects of null alleles for the Bmp1, Tll1, and Tll2 genes are limited to individual tissues, should enable survival of mice to developmental stages that will allow testing of which BMP-1-like proteinases or combination of proteinases may be involved in cleaving procollagen VII in skin.

Type VII procollagen is produced by keratinocytes, whereas it has been suggested that the majority of type VII procollagen-cleaving activity is provided by fibroblasts at the dermal-epidermal junction [191]. The latter is in contrast to indications that laminin-5-processing activity is provided by keratinocytes [120]. Clearly, the full range of roles of cells and proteinases involved in forming the mature laminin-5 and type VII collagen of the dermal-epidermal basement membrane zone remains to be fully elucidated.

5
Processing/Shedding of Cell Surface Collagen Types XIII, XVII and XXV

Certain collagens are not exclusively ECM components, but rather spend some portion of their extracellular existences as integral proteins of the plasma membrane. Three such collagens are types XIII, XVII and XXV. Although these three collagen types do not share high degrees of sequence homology, each is a type II transmembrane protein with an NH_2-terminal cytoplasmic domain, a hydrophobic transmembrane domain located in the NH_2-terminal portion of

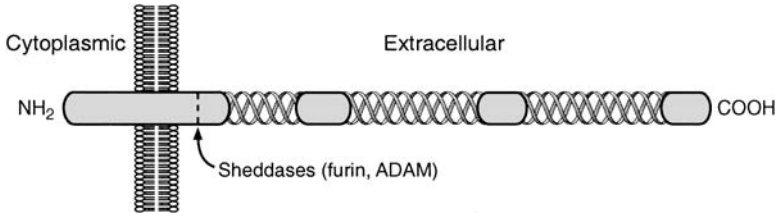

Fig. 7 Processing of cell surface collagens

the molecule, and a relatively large extracellular domain [193–195], that is proteolytically cleaved within sequences closely juxtaposed to the plasma membrane (Fig. 7).

A small proportion of type XIII collagen, which has been suggested to play roles in binding soluble ligands and/or ECM components to cells [193], is cleaved from the surfaces of cultured cells and is found in culture media [196]. Recombinant forms of XIII collagen lacking the small NH_2-terminal cytoplasmic domain are much more susceptible to cleavage than are full-length forms, suggesting some level of control by this domain over processing of extracellular sequences [197]. NH_2-terminal sequencing of the shed ectodomain shows that cleavage occurs immediately COOH-terminal to a consensus recognition site for cleavage by furin-like proprotein convertases, and furin-specific inhibitor decanoyl-RVKR-chloromethyl ketone inhibits ectodomain shedding [197]. Thus, some portion of type XIII collagen molecules, found on the surface of a variety of cell types, seems to be shed via proteolytic cleavage by furin-like proteinase(s), although the biological significance of this process is at present unknown.

Collagen XXV is expressed by neurons of the brain, and its shed ectodomain binds to fibrils of amyloid β peptide, with which it is co-deposited in senile plaques associated with Alzheimer's disease [195]. Collagen XXV ectodomain is apparently shed via cleavage by furin-like proteinases, since amino acid substitutions within a consensus sequence for cleavage by furin-like proteinases inhibits cleavage [195].

Type XVII collagen, also known as the 180-kDa bullous pemphigoid antigen (BP180), is an integral component of epithelial cell hemidesmosomes, involved in securing cells to underlying basement membranes [198]. The importance of type XVII collagen is underscored by phenotypes of the disease junctional epidermolysis bullosa and autoimmune diseases of the pemphigoid group, in which mutations in the type XVII collagen gene and autoantibodies against type XVII collagen, respectively, lead to decreased epidermal adhesion and severe blistering [199, 200]. The ectodomain of collagen XVII is proteolytically shed, and such shedding is inhibited by incubating keratinocytes in the presence of decanoyl-RVKR-chloromethyl ketone [201]. However, furin does not cleave purified type XVII collagen in vitro, whereas shedding is potentiated by culturing cells in the presence of phorbol esters and is inhibited by a specific

spectrum of metalloproteinase inhibitors, both of which effects fit the profile of either MMPs or ADAM family metalloproteinases [202]. Cells from knockout mice null for candidate MMPs, such as MT1-MMP and MMP-2, show no diminishment in collagen XVII shedding [202]. Furthermore, cleavage of collagen XVII by MMPs produces fragments that differ in size from those produced via physiological processing, and a specific gelatinase inhibitor that strongly inhibits the activities of candidate MMPs MMP-2 and MMP-9 does not inhibit collagen XVII shedding [202]. In contrast, the ADAM family member TACE cleaves type XVII collagen to produce fragments identical in size to those produced by cells, and keratinocytes from TACE knockout mice show a marked decrease in collagen XVII shedding [202]. Thus ADAM family members TACE, ADAM-9 and ADAM-10, all of which are expressed in keratinocytes [202], are candidates for involvement in in vivo processing of collagen XVII. Presumably, the ability of decanoyl-RVKR-chloromethyl ketone to inhibit collagen XVII shedding is indirect, since furin-like activity is necessary for activation of ADAM family members [203]. Further information concerning attributes of the ADAM membrane-anchored metalloprotease sheddases, which are involved in myriad processes in development and homeostasis, can be found in the review by Becherer and Blobel [203]. Although the physiological roles of collagen shedding are unclear, the shed ectodomain of type XVII collagen is bound by keratinocytes, and processing of collagen XVII may contribute to altered motility of epithelial cells [202].

A number of additional type II integral membrane proteins exist that contain collagenous regions in their ectodomains [204]. One of these, collagen XXIII, is known only via data mining of the genome databases, and whether this collagen is proteolytically processed remains to be determined. Remaining transmembrane proteins with collagenous domains are not formally collagens [204]. Only one of these proteins, ectodysplasin A (EDA) is known, at this time, to be proteolytically processed. Ectodomains of alternatively spliced versions of EDA play roles in epidermal development [205] and mutations in the gene that encodes EDA result in the genetic condition X-linked anhidrotic/hypohidrotic ectodermal dysplasia, characterized by impaired development of hair, sweat glands and teeth [206]. Proteolytic shedding of EDA appears to be via furin-like convertase activity, and mutations that cause single amino acid substitutions within the furin site consensus sequence block shedding of the ectodomain and cause hypohidrotic ectodermal dysplasia [207]. This latter effect shows processing of the ectodomain to be important to the developmental roles of EDA [207].

6
Processing of Multiplexin Collagen Types XV and XVIII

The two nonfibrillar collagen types XV and XVIII have similar domain structures and together have been referred to as the multiplexin (multiple triplehelix domains and interruptions) collagen subfamily [208]. Collagens XV and

XVIII are heparan sulfate [209] and chondroitin sulfate [210] proteoglycans, respectively, and both are basement membrane components [211–213]. Functions of intact collagens XV and XVIII are not well understood, although mice null for the collagen XV gene have mild, progressive degeneration of skeletal muscle, susceptibility to exercise-induced muscle damage, and abnormal microvasculature of the heart and skeletal muscle [214]; and mice null for the collagen XVIII gene show subtle defects in the iris and microvasculature of the eye [215, 216]. Furthermore, mutations in the human gene for collagen XVIII can result in Knobloch syndrome, which is characterized by various eye abnormalities [217]. Of particular interest is the finding that a cleaved subdomain, designated endostatin, of the COOH-terminal noncollagenous NC1 domain of collagen XVIII, can have potent antiangiogenic effects in certain in vivo settings [218]. Although such effects have been demonstrated via administration of exogenous recombinant endostatin [218], relatively high levels of endogenous endostatin are detected in mouse and human sera [219], suggesting the possibility of physiological roles. However, endogenous endostatin of normal serum exhibits heterogeneous NH_2-termini, which differ from the single NH_2-terminus of the endostatin originally employed in angiogenesis inhibition assays and derived from conditioned media of the EOMA line of murine hemangioendothelioma cells [218]. In fact, endostatin, a compact autonomously folding proteinase-resistant domain, is linked to the remainder of type XVIII via a proteinase-sensitive hinge region, within which cleavage appears to occur via multiple proteolytic pathways [219]. Several data suggest that cleavage may be a two step process with initial cleavage of the entire NC1 domain trimer by metalloproteinase(s), and subsequent cleavage within the proteinases-sensitive hinge region to release endostatin [220, 221]. Cleaved, intact NC1 domain may be retained in basement membranes due to strong non-covalent interactions with matrix components, whereas subsequent cleavage may result in the diffusion of the isolated endostatin domain, which does not bind as strongly to matrix components [219]. Serine elastases are capable of cleaving at the same site at which endostatin is cleaved in EOMA cell cultures, and administering elastase inhibitors, such as elastatinal, to EOMA cells in one study inhibited the production of endostatin, with a concomitant accumulation of intact NC1 domains [220]. Curiously, another study showed that treatment of EOMA cells with elastatinal did not inhibit production of endostatin [221]. This same study showed cathepsin L to cleave endostatin at the same site employed in EOMA cell cultures, showed EOMA cells to secrete cathepsin L, and showed that an inhibitor specific to cathepsins inhibits proteolytic production of endostatin in EOMA cell cultures [221]. Thus cathepsin L, which has an acidic pH optimum, has been suggested to cleave endostatin in the acidic pericellular microenvironment of tumors [221]. However, different tissues and sera contain species of endostatin that vary in their NH_2-termini, suggesting that various proteinases are involved in processing endostatin in vivo [219]. In a survey, 11 of 12 proteinases tested cleaved within the same 15 residue span of proteinase-sensitive hinge region as that in which in vivo cleavages oc-

cur, and some of the same proteinases were capable of degrading endostatin upon longer incubations [222]. Thus, it remains to be determined which proteinase(s) may be involved in cleaving collagen XVIII in vivo to produce physiologically functional forms of endostatin, and which are involved in catabolic degradation of collagen XVIII or artifactual processing associated with isolation of endostatin from tissues and culture media. Fragments corresponding to the endostatin domain of collagen XV have been isolated from human blood filtrate [223] and detected in mouse tissues [224], suggesting possible physiological significance. However, although the collagen XV endostatin-like fragment has been shown to have anti-angiogenic effects in assays [224, 225], the nature of the proteolytic processing of this fragment has not been explored at this time.

7
Concluding Remarks

Processing of the major and minor fibrillar procollagens by mammalian BMP-1-like proteinases likely coordinates fibrillogenesis with other biosynthetic events in vivo, since the same proteinases are responsible for the processing of other structural molecules and for regulating signaling by certain TGF-β-related growth factors and morphogens. Uncovering the full range of processes to which fibrillogenesis is linked will require identification of additional in vivo substrates of the BMP-1-like proteinases. Such identification will require, in addition to the candidate substrate approach, high throughput methodologies, such as mass-spectrometric analysis of cleavage products produced by wild type cells, but not by the cells of mice null for various combinations of BMP-1-like proteinase genes. High throughput screens may also be of use in comparison of cleavage products in untreated normal cell cultures and cultures of the same cells treated with small molecule inhibitors reportedly specific for the mammalian BMP-1-like proteinases [120, 226].

Clearly, the same types of high throughput methodologies can be conducted for the pNPs, using knockout mice null for various combinations of the genes for ADAMTS-2, -3 and -14, such that functional redundancies will have been removed, or using highly specific inhibitors, should they become available. Additional high throughput screens, such as phage display with cDNA expression libraries representing various tissues, should also identify novel binding partners of the BMP-1-like and ADAMTS proteinases. Such binding partners could include novel substrates, and possible endogenous inhibitors and enhancers of proteinase activity. Another future goal is to define physiological roles for cleavage/shedding of cell surface collagens, perhaps via use of knockin mutations that destroy relevant cleavage sites. Clearly, further understanding of the processing of collagens is important, as such processing represents important control points in the regulation of collagen functions and in the integration of collagen biosynthesis with other in vivo events.

References

1. Sternlicht MD, Werb Z (2001) Annu Rev Cell Dev Biol 17:463
2. Brinckerhoff CE, Matrisian LM (2002) Nat Rev Mol Cell Biol 3:207
3. Lauer-Fields JL, Juska D, Fields GB (2002) Biopolymers 66:19
4. For reviews: (a) Kuhn K (1987) The classical collagens: types I, II and III. In: Mayne R, Burgeson RE (eds) Structure and function of collagen types. Academic Press, Orlando, p 1; (b) van der Rest M, Garrone R (1991) FASEB J 5:2814; (c) Prockop DJ, Kivirikko KI (1995) Annu Rev Biochem 64:403; (d) Myllyharju J, Kivirikko KI (2001) Ann Med 33:7
5. Cole WG, Evans R, Sillence DO (1987) J Med Genet 24:698
6. Byers PH, Duvic M, Atkinson M, Robinow M, Smith LT, Krane SM, Greally MT, Ludman M, Matalon R, Pauker S, Quanbeck D, Schwarze U (1997) Am J Med Genet 72:94
7. Byers PH (1995) Disorders of collagen biosynthesis and structure. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic and molecular bases of inherited disease. McGraw-Hill, New York, p 4029
8. Prockop DJ, Hulmes DJS (1994) Assembly of collagen fibrils de novo from soluble precursors: polymerization and copolymerization of procollagen, pN-collagen, and mutated collagens. In: Yurchenco PD, Birk DE, Mechan RP (eds) Extracellular matrix assembly and structure. Academic Press, New York, p 47
9. Hojima Y, McKenzie J, van der Rest M, Prockop DJ (1989) J Biol Chem 264:11336
10. Hojima Y, Mörgelin MM, Engel J, Boutillon MM, van der Rest M, McKenzie J, Chen G-C, Rafi N, Romani AM, Prockop DJ (1994) J Biol Chem 269:11381
11. Hojima Y, van der Rest M, Prockop DJ (1985) J Biol Chem 260:15996
12. Kessler E, Adar R, Goldberg B, Niece R (1986) Collagen Relat Res 6:249
13. Kessler E, Takahara K, Biniaminov L, Brusel M, Greenspan DS (1996) Science 271:360
14. Lapière CM, Lenaers A, Kohn LD (1971) Proc Natl Acad Sci USA 68:3054
15. Tuderman L, Kivirikko KI, Prockop DJ (1978) Biochemistry 17:2948
16. Tuderman L, Prockop DJ (1982) Eur J Biochem 125:545
17. Halila R, Peltonen L (1984) Biochemistry 23:1251
18. Halila R, Peltonen L (1986) Biochem J 239:47
19. Nusgens BV, Goebels Y, Shinkai H, Lapiere CM (1980) Biochem J 191:699
20. Colige A, Beshin A, Samyn B, Goebels Y, Van Beeumen J, Nusgens BV, Lapière CM (1995) J Biol Chem 270:16724
21. Colige A, Li S-W, Sieron AL, Nusgens BV, Prockop DJ, Lapière CM (1997) Proc Natl Acad Sci USA 94:2374
22. Colige A, Sieron AL, Li S-W, Schwarze U, Petty E, Wertelecki W, Wicox W, Krakow D, Cohn DH, Reardon W, Byers PH, Lapière CM, Prockop DJ, Nusgens BV (1999) Am J Hum Genet 65:308
23. Kuno, K, Kanada N, Nakashima E, Fujiki F, Ichimura F, Matsushima K (1997) J Biol Chem 272:556
24. Cal S, Obaya AJ, Llamazares M, Garabaya C, Quesada V Lopez-Otin C (2002) Gene (Amst) 283:49
25. Somerville RP, Longpre JM, Engle JM, Jungers KA, Ross M, Evanko S, Wight TN, Leduc R, Apte SS (2003) J Biol Chem 278:9503
26. Hurskainen TL, Hirohata S, Seldin MF, Apte SS (1999) J Biol Chem 274:25555
27. Kuno K, Matsushima K (1998) J Biol Chem 273:13912
28. Tortorella MD, Burn TC, Pratta MA, Abbaszade I, Hollis JM, Liu R, Rosenfeld SA, Copeland RA, Decicco CP, Wynn R, Rockwell A, Yang F, Duke JL, Solomon K, George H, Bruckner R, Nagase H, Itoh Y, Ellis DM, Ross H, Wiswall BH, Murphy K, Hillman MC Jr, Hollis GF, Newton RC, Magolda RL, Trzaskos JM, Arner EC (1999) Science 284:1664

29. Abbaszade, I, Liu R-Z, Yang F, Rosenfeld SA, Ross OH, Link JR, Ellis DM, Tortorella MD, Pratta MA, Hollis JM, Wynn R, Duke JL, George HJ, Hillman MC Jr, Murphy K, Wiswall BH, Copeland RA, Decicco CP, Bruckner R, Nagase H, Itoh Y, Newton RC, Magolda RL, Trzaskos JM, Hollis GF, Arner EC, Burn TC (1999) J Biol Chem 274:23443
30. Kuno K, Okada Y, Kawashima H, Nakamura H, Miyasaka M, Ohno H, Matsushima K (2000) FEBS Lett 478:241
31. Shindo T, Kurihara H, Kuno K, Yokoyama H, Wada T, Kurihara Y, Imai T, Wang Y, Ogata M, Nishimatsu H, Moriyama N, Oh-hashi Y, Morita H, Ishikawa T, Nagai R, Yazaki Y, Matsushima K (2000) J Clin Invest 105:1345
32. Vázquez F, Hastings G, Ortega M-A, Lane TF, Oikemus S, Lombardo M, Iruela-Arispe ML (1999) J Biol Chem 274:23349
33. Luque A, Carpizo DR, Iruela-Arispe ML (2003) J Biol Chem 278:23656
34. Blelloch R, Kimble J (1999) Nature 399:586
35. Levy, GG, Nichols WC, Lian EC, Foroud T, McClintick JN, McGee BM, Yang AY, Siemlenlak DR, Stark KR, Grupppo R, Sarode R, Shurin SB, Chandrasekaran V, Stabler SP, Sablo H, Bouhassira EE, Upshaw JD Jr, Ginsburg D, Tsai H-M (2001) Nature 413:488
36. Li S-W, Arita M, Fertala A, Bao Y, Kopen GC, Långsjö TK, Hyttinen MM, Helminen HJ, Prockop DJ (2001) Biochem J 355:271
37. Fernandes RJ, Hirohata S, Engle JM, Colige A, Cohn DH, Eyre DR, Apte SS (2001) J Biol Chem 276:31502
38. Schnieke A, Harbers K, Jaenisch R (1983) Nature 304:315
39. Wang W-M, Lee S, Steiglitz BM, Scott IC, Lebares CC, Allen ML, Brenner MC, Takahara K, Greenspan DS (2003) J Biol Chem 278:19549
40. Colige A, Vandenberghe I, Thiry M, Lambert CA, Van Beeumen J, Li S-W, Prockop DJ, Lapière CM, Nusgens BV (2002) J Biol Chem 277:5756
41. Kessler E, Adar R (1989) Eur J Biochem 186:115
42. Li S-W, Sieron AL, Fertala A, Hojima Y, Arnold WV, Prockop DJ (1996) Proc Natl Acad Sci USA 93:5127
43. Wozney JM, Rosen V, Celeste AJ, Misock LM, Whitters MJ, Kriz RW, Hewick RM, Wang EA (1988) Science 242:1528
44. Wang EA, Rosen V, Cordes P, Hewick RM, Kriz MJ, Luxenberg DP, Sibley BS, Wozney JM (1988) Proc Natl Acad Sci USA 85:9484
45. Celeste AJ, Iannazzi JA, Taylor RC, Hewick RM, Rosen V, Wang EA, Wozney JM (1990) Proc Natl Acad Sci USA 87:9843
46. Bond JS, Beynon RJ (1995) Protein Sci 4:1247
47. Leighton M, Kadler KE (2003) J Biol Chem 278:18478
48. Tosi M, Duponchel C, Meo T, Julier C (1987) Biochemistry 26:8516
49. Bork P, Beckmann G (1993) J Mol Biol 231:539
50. Handford PA, Baron M, Mayhew M, Willis A, Beesley T, Brownlee GG, Campbell ID (1990) EMBO J 9:475
51. Hartigan N, Garrigue-Antar L, Kadler KE (2003) J Biol Chem 278:18045
52. Shimell MJ, Ferguson EL, Childs SR, O'Connor MB (1991) Cell 67:469
53. Childs SB, O'Connor MB (1994) Dev Biol 162:209
54. Finelli AL, Bossie CA, Xie T, Padgett RW (1994) Development 120:861
55. Padgett RW, Wozney JM, Gelbart WM (1993) Proc Natl Acad Sci USA 90:2905
56. Ducy P, Karsenty G (2000) Kid Int 57:2207
57. Marqués G, Musacchio M, Shimell MJ, Wünnenberg-Stapleton K, Cho KW, O'Connor MB (1997) Cell 91:417
58. (a) Piccolo S, Agius E, Lu B, Goodman S, Dale L, De Robertis EM (1997) Cell 91:407; (b) Blader P, Rastegar S, Fischer N, Strähle U (1997) Science 278:1937

59. Takahara K, Lyons GE, Greenspan DS (1994) J Biol Chem 269:32572
60. Takahara K, Brevard R, Hoffman GG, Suzuki N, Greenspan DS (1996) Genomics 34:157
61. Scott IC, Blitz IL, Pappano WN, Imamura Y, Clark TG, Steiglitz BM, Thomas CL, Maas SA, Takahara K, Cho KWY, Greenspan DS (1999) Dev Biol 213:283
62. Pappanno WN, Steiglitz BM, Scott IC, Keene DR, Greenspan DS (2003) Mol Cell Biol 23:4428
63. Suzuki N, Labosky PA, Furuta Y, Hargett L, Dunn R, Fogo AB, Takahara K, Peters DMP, Greenspan DS, Hogan BLM (1996) Development 122:3587
64. Scott IC, Imamura Y, Pappano WN, Troedel JM, Recklies AD, Roughley PJ, Greenspan DS (2000) J Biol Chem 275:30504
65. Lee S, Solow-Cordero DE, Kessler E, Takahara K, Greenspan DS (1997) J Biol Chem 272:19059
66. Trelstad RL, Hayashi K (1979) Dev Biol 71:228
67. Birk DE, Trelstad RL (1984) J Cell Biol 99:2024
68. Bonfanti L, Mironov AA Jr, Martínez-Menárguez JA, Martella O, Fusella A, Baldassarre M, Buccione R, Geuze HJ, Mironov AA, Luini A (1998) Cell 95:993
69. Hojima Y, Behta B, Romanic AM, Prockop DJ (1994) Anal Biochem 223:173
70. Morris NP, Fesslear LI, Weinstock A, Fessler JH (1975) J Biol Chem 250:5719
71. Uitto J, Lichtenstein JR (1976) Biochem Biophys Res Commun 71:60
72. Adar R, Kessler E, Goldberg B (1986) Collagen Relat Res 6:267
73. Takahara K, Kessler E, Biniaminov L, Brusel M, Eddy RL, Jani-Sait S, Shows TB, Greenspan DS (1994) J Biol Chem 269:26280
74. Bányai L, Patthy L (1999) Protein Sci 8:1636
75. Mott JD, Thomas CL, Rosenbach MT, Takahara K, Geenspan DS, Banda MJ (2000) J Biol Chem 275:1384
76. Masuda M, Igarasi H, Kano M, Yoshikura H (1998) Cell Growth Differ 9:381
77. Ricard-Blum S, Bernocco S, Font B, Moali C, Eichenberger D, Farjanel J, Burchardt ER, van der Rest M, Kessler E, Hulmes DJS (2002) J Biol Chem 277:33864
78. Xu H, Acott TS, Wirtz MK (2000) Genomics 66:264
79. Steiglitz BM, Keene DR, Greenspan DS (2002) J Biol Chem 277:49820
80. Iozzo RV (1998) Annu Rev Biochem 67:609
81. Xu T, Bianco P, Fisher LW, Longenecker G, Smith E, Goldstein S, Bonadio J, Boskey A, Heegaard A-M, Sommer B, Satomura K, Dominguez P, Chengyan Z, Kulkarni AB, Robey PG, Young MF (1998) Nat Genet 20:78
82. Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, Iozzo RV (1997) J Cell Biol 136:729
83. Bianco P, Fisher LW, Young MF, Termine JD, Robey PG (1990) J Histochem Cytochem 38:1549
84. Scott JE, Orford CR (1981) Biochem J 197:213
85. Brown DC, Vogel KG (1989) Matrix 9:468
86. Schönherr E, Witsch-Prehm P, Harrach B, Robenek H, Rauterberg J, Kresse H (1995) J Biol Chem 270:2776
87. Bidanset DJ, Guidry C, Rosenberg LC, Choi HU, Timpl R, Höök M (1992) J Biol Chem 267:5250
88. Schmidt G, Hausser H, Kresse H (1991) Biochem J 280:411
89. Hildebrand A, Romaris M, Rasmussen LM, Heinegård D, Twardzik DR, Borders AW, Ruoslahti E (1994) Biochem J 302:527
90. Yamaguchi Y, Mann DM, Ruoslahti E (1990) Nature 346:281
91. Roughley PJ, White RJ, Mort JS (1996) Biochem J 318:779
92. Fisher LW, Termine JD, Young MF (1989) J Biol Chem 264:4571

93. Krusius T, Ruoslahti E (1986) Proc Natl Acad Sci USA 83:7683
94. Fisher LW, Hawkins GR, Tuross N, Termine JD (1987) J Biol Chem 262:9702
95. Choi HU, Johnson TL, Pal S, Tang L-H, Rosenberg L, Neame PJ (1989) J Biol Chem 264:2876
96. Roughley PJ, White RJ (1989) Biochem J 262:823
97. Svensson L, Oldberg Å, Heinegård D (2001) Osteoarthritis Cartilage 9:S23
98. Funderburgh JL, Corpuz LM, Roth MR, Funderburgh ML, Tasheva ES, Conrad GW (1997) J Biol Chem 272:28089
99. Johnson HJ, Rosenberg L, Choi HU, Garza S, Höök M, Neame PJ (1997) J Biol Chem 272:18709
100. Ge G, Seo N-S, Höök M, Greenspan DS (2004) J Biol Chem 279: 41626
101. Tasheva ES, Koester A, Paulsen AQ, Garrett AS, Boyle DL, Davidson HJ, Song M, Fox N, Conrad GW (2002) Mol Vis 8:407
102. Kagan HM, Trackman PC (1991) Am J Respir Cell Mol Biol 5:206
103. Trackman PC, Bedell-Hogan D, Tang J, Kagan HM (1992) J Biol Chem 267:8666
104. Panchenko MV, Stetler-Stevenson WG, Trubetskoy OV, Gacheru SN, Kagan HM (1996) J Biol Chem 271:7113
105. Uzel MI, Scott IC, Babakhanlou-Chase H, Palamakumbura AH, Pappano WN, Hong H-H, Greenspan DS, Trackman PC (2001) J Biol Chem 276:22537
106. Butler WT (1998) Eur J Oral Sci 106, Suppl 1:204
107. Veis A (1993) J Bone Miner Res 8, Suppl 2:493
108. Zhang X, Zhao J, Li C, Gao S, Qiu C, Liu P, Wu G, Qiang B, Lo WH, Shen Y (2001) Nat Genet 27:151
109. Thotakura SR, Mah T, Srinivasan R, Takagi Y, Veis A, George A (2000) J Dent Res 79:835
110. Sreenath T, Thyagarajan T, Hall B, Longenecker G, D'Souza R, Hong S, Wright JT, MacDougall M, Sauk J, Kulkarni AB (2003) J Biol Chem 278:24874
111. Qin C, Cook RG, Orkiszewski RS, Butler WT (2001) J Biol Chem 276:904
112. Qin C, Brunn JC, Cook RG, Orkiszewski RS, Malone JP, Veis A, Butler WT (2003) J Biol Chem 278:34700
113. Hunter GK, Hauschka PV, Poole AR, Rosenberg LC, Goldberg HA (1996) Biochem J 317:59
114. He G, Dahl T, Veis A, George A (2003) Nat Mater 2, 552
115. Steiglitz BM, Ayala M, Narayanan K, George A, Greenspan DS (2004) J Biol Chem 279:980
116. Rousselle P, Lunstrum GP, Keene DR, Burgeson RE (1991) J Cell Biol 114:567
117. Christiano AM, Uitto J (1996) Exp Dermatol 5:1
118. Marinkovitch MP, Lunstrum GP, Burgeson RE (1992) J Biol Chem 267:17900
119. Giannelli G, Falk-Marzillier J, Schiraldi O, Stetler-Stevenson WG, Quaranta V (1997) Science 277:225
120. Veitch DP, Nokelainen P, McGowan KA, Nguyen T-T, Nguyen NE, Stephenson R, Pappano WN, Keene DR, Spong SM, Greenspan DS, Findell PR, Marinkovich MP (2003) J Biol Chem 278:15661
121. Koshikawa N, Giannelli G, Cirulli V, Miyazaki K, Quaranta V (2000) J Cell Biol 148:615
122. Amano S, Scott IC, Takahara K, Koch M, Champliaud M-F, Gerecke DR, Keene DR, Hudson DL, Nishiyama T, Lee S, Greenspan DS, Burgeson RE (2000) J Biol Chem 275:22728
123. Goldfinger LE, Stack MS, Jones JC (1998) J Cell Biol 141:255
124. Lee S-J, McPherron AC (2001) Proc Natl Acad Sci USA 98:9306
125. Hill JJ, Davies MV, Pearson AA, Wang JH, Hewick RM, Wolfman NM, Qiu Y (2002) J Biol Chem 277:40735
126. Zimmers TA, Davies MV, Koniaris LG, Haynes P, Esquela AF, Tomkinson KN, McPherron AC, Wolfman NM, Lee S-J (2002) Science 296:1486

127. McPherron AC, Lawler AM, Lee S-J (1997) Nature 387:83
128. Thies RS, Chen T, Davies MV, Tomkinson KN, Pearson AA, Shakey QA, Wolfman NM (2001) Growth Factors 18:251
129. Wolfman NM, McPherron AC, Pappano WN, Davies MV, Song K, Tomkinson KN, Wright JF, Zhao L, Sebald SM, Greenspan DS, Lee S-J (2003) Proc Natl Acad Sci USA 100:15842
130. Birk DE, Fitch JM, Babiarz JP, Linsenmayer TG (1988) J Cell Biol 106:999
131. Birk DE, Fitch JM, Babiarz JP, Doanne KJ, Linsenmayer TF (1990) J Cell Sci 95:649
132. Nicholls AC, Olliver JE, McCarron S, Harrison JB, Greenspan DS, Pope FM (1996) J Med Genet 33:940
133. Toriello HV, Glover TW, Takahara K, Byers PH, Miller DE, Higgins JV, Greenspan DS (1996) Nat Genet 13:361
134. Wenstrup RJ, Langland GT, Willing MC, D'Soua VN, Cole WG (1996) Num Mol Genet 5:1733
135. Mendler M, Eich-Bender SG, Vaughn L, Winterhalter KH, Bruckner P (1989) J Cell Biol 108:191
136. Li Y, Lacerda DA, Warman ML, Beier DR, Yoshioka H, Ramirez F, Wardell BB, Lifferth GD, Teuscher C, Woodward SR, Taylor BA, Seegmiller RE, Olsen BR (1995) Cell 80:423
137. Fessler JH, Fessler LI (1987) Type V collagen. In: Mayne R, Burgeson RE (eds) Structure and function of collagen types. Academic Press, Orlando, p 81
138. Haralson MA, Mitchell WM, Rhodes RK, Kresina TF, Gay R, Miller EJ (1980) Proc Natl Acad Sci USA 77:5206
139. Kumamoto CA, Fessler JH (1981) J Biol Chem 256:7053
140. Moradi-Améli M, Rousseau J-C, Kleman J-P, Champliaud M-F, Boutillon M-M, Bernillon J, Wallach J, van der Rest M (1994) Eur J Biochem 221:987
141. Sage H, Bornstein P (1979) Biochemistry 18:3815
142. Rhodes RK, Miller EJ (1981) Collagen Relat Res 1:337
143. Abedin MZ, Ayad S, Weiss JB (1982) Biosci Rep 2:493
144. van der Rest M, Garrone R (1991) FASEB J 5:2814
145. Brown RA, Shuttleworth CA, Weiss JB (1978) Biochem Biophys Res Commun 80:866
146. Imamura Y, Scott IC, Greenspan DS (2000) J Biol Chem 275:8749
147. Chernousov MA, Rothblum K, Tyler WA, Stahl RC, Carey DJ (2000) J Biol Chem 275: 28208
148. Morris NP, Bächinger HP (1987) J Biol Chem 262:11345
149. Niyibizi C, Eyre DR (1989) FEBS Lett 242:314
150. Eyre D, Wu J-J (1987) Type XI collagen. In: Mayne R, Burgeson RE (eds) Structure and function of collagen types. Academic Press, Orlando, p 261
151. Kleman J-P, Hartmann DJ, Ramirez F, van der Rest M (1992) Eur J Biochem 210:329
152. Mayne R, Brewton RG, Mayne PM, Baker JR (1993) J Biol Chem 268:9381
153. Broek DL, Madri J, Eikenberry EF, Brodsky B (1985) J Biol Chem 260:555
154. Thom JR, Morris NP (1991) J Biol Chem 266:7262
155. Linsenmayer TF, Gibney E, Igoe F, Gordon MK, Fitch JM, Fessler LI, Birk DE (1993) J Cell Biol 121:1181
156. Rousseau J-C, Farjanel J, Boutillon M-M, Hartmann DJ, van der Rest M, Moradi-Améli M (1996) J Biol Chem 271:23743
157. Niyibizi C, Eyre DR (1993) Biochim Biophys Acta 1203:304
158. Greenspan DS, Cheng W, Hoffman GG (1991) J Biol Chem 266:24727
159. Takahara K, Sato Y, Okazawa K, Okamoto N, Noda A, Yaoi Y, Kato I (1991) J Biol Chem 266:13124

160. Bernard M, Yoshioka H, Rodriguez E, van der Rest M, Kimura T, Ninomiya Y, Olsen BR, Ramirez F (1988) J Biol Chem 263:17159
161. Kimura T, Cheah KSE, Chan SDH, Lui VCH, Mattei M-G, van der Rest M, Ono K, Solomon E, Ninomiya Y, Olsen BR (1989) J Biol Chem 264:13910
162. Tsumaki N, Kimura T (1995) J Biol Chem 270:2372
163. Zhidkova NI, Brewton RG, Mayne R (1993) FEBS Lett 326:25
164. Yoshioka H, Ramirez F (1990) J Biol Chem 265:6423
165. Takahara K, Hoffman GG, Greenspan DS (1995) Genomics 29:588
166. Vuoristo MM, Pihlajamaa T, Vandenberg P, Prockop DJ, Ala-Kokko L (1995) J Biol Chem 270:22873
167. Weil D, Bernard M, Gargano S, Ramirez F (1987) Nucleic Acids Res 15:181
168. Woodbury D, Benson-Chanda V, Ramirez F (1989) J Biol Chem 264:2735
169. Imamura Y, Steiglitz BM, Greenspan DS (1998) J Biol Chem 273:27511
170. Kessler E, Fichard A, Chanut-Delalande H, Brusel M, Ruggiero F (2001) J Biol Chem 276:27051
171. Unsöld C, Pappano WN, Imamura Y, Steiglitz BM, Greenspan DS (2002) J Biol Chem 277:5596
172. Lee S-T, Kessler E, Greenspan DS (1990) J Biol Chem 265:21992
173. Medeck RJ, Sosa S, Morris N, Oxford JT (2003) Biochem J 376:361
174. Oxford JT, Doege KJ, Morris NP (1995) J Biol Chem 270:9478
175. Zhidkova NI, Justice SK, Mayne R (1995) J Biol Chem 270:9486
176. Steiner DF (1998) Curr Opin Chem Biol 2:31
177. Nakayama K (1997) Biochem J 327:625
178. Denault JB, Leduc R (1996) FEBS Lett 379:113
179. Koch M, Laub F, Zhou P, Hahn RA, Tanaka S, Burgeson RE, Gerecke DR, Ramirez F, Gordon MK (2003) J Biol Chem 278:43236
180. Pace JM, Corrado M, Missero C, Byers PH (2003) Matrix Biol 22:3
181. Sakai LY, Keene DR, Morris NP, Burgeson RE (1986) J Cell Biol 103:1577
182. Lunstrum GP, Sakai LY, Keene DR, Morris NP, Burgeson RE (1986) J Biol Chem 261:9042
183. Morris NP, Keene DR, Glanville RW, Bentz H, Burgeson RE (1986) J Biol Chem 261: 5638
184. Keene DR, Sakai LY, Lunstrum GP, Morris NP, Burgeson RE (1987) J Cell Biol 104:611
185. Christiano AM, Greenspan DS, Hoffman GG, Zhang X, Tamai Y, Lin AN, Dietz HC, Hovnanian A, Uitto J (1993) Nat Genet 4:62
186. Bruckner-Tuderman L (1999) J Dermatol Sci 20:122
187. Bruckner-Tuderman L (1999) Matrix Biol 18:43
188. Chen M, Keene DR, Costa FK, Tahk SH, Woodley DT (2001) J Biol Chem 276:21649
189. Lunstrum GP, Kuo H-J, Rosenbaum LM, Deene DR, Glanville RW, Sakai LY, Burgeson RE (1987) J Biol Chem 262:13706
190. Bruckner-Tuderman L, Nilssen Ø, Zimmermann DR, Dours-Zimmermann MT, Kalinke DU, Gedde-Dahl T Jr, Winberg J-O (1995) J Cell Biol 131:551
191. Rattenholl A, Pappano WN, Koch M, Keene DR, Kadler KE, Sasaki T, Timple R, Burgeson RE, Greenspan DS, Bruckner-Tuderman L (2002) J Biol Chem 277:26372
192. Clark TG, Conway SJ, Scott IC, Labosky PA, Winnier G, Bundy J, Hogan BLM, Greenspan DS (1999) Development 126:2631
193. Hägg P, Rehn M, Huhtala P, Väisänen T, Tamminen M, Pihlajaniemi T (1998) J Biol Chem 273:15590
194. Giudice GJ, Emery DJ, Dia LA (1992) J Invest Dermatol 99:243
195. Hashimoto T, Wakabayashi T, Watanabe A, Kowa H, Hosoda R, Nakamura A, Kanazawa I, Arai T, Takio K, Mann DMA, Iwatsubo T (2002) EMBO J 21:1524

196. Peltonen S, Hentula M, Hägg P, Ylä-Outinen H, Tuukkanen J, Lakkakorpi J, Rehn M, Pihlajaniemi T, Peltonen J (1999) J Invest Dermatol 113:635
197. Snellman A, Tu H, Väisänen T, Kvist A-P, Huhtala P, Pihlajaniemi T (2000) EMBO J 19:5051
198. Borradori L, Sonnenberg A (1999) J Invest Dermatol 112:411
199. McGrath JA, Gatalica B, Christiano AM, Li K, Owaribe K, McMillan JR, Eady RA, Uitto J (1995) Nat Genet 11:83
200. Schumann H, Baetge J, Tasanen K, Wojnarowska F, Schacke H, Zillikens D, Bruckner-Tuderman L (2000) Am J Pathol 156:685
201. Schäcke H, Schumann H, Hammami-Hauasli N, Raghunath M, Bruckner-Tuderman L (1998) J Biol Chem 273:25937
202. Franzke C-W, Tasanen K, Schäcke H, Zhou Z, Tryggvason K, Mauch C, Zigrino P, Sunnarborg S, Lee DC, Fahrenholz F, Bruckner-Tuderman L (2002) EMBO J 21:5026
203. Becherer JD, Blobel CP (2003) Curr Top Dev Biol 54:101
204. Franzke C-W, Tasanen K, Schumann H, Bruckner-Tuderman L (2003) Matrix Biol 22:299
205. Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, Goddard A, de Vos AM, Gao W-Q, Dixit VM (2000) Science 290:523
206. Kere J, Srivastava AK, Montonen O, Zonana J, Thomas NST, Ferguson B, Munoz F, Morgan D, Clarke A, Baybayan P, Chen EY, Ezer S, Saarialho-Kere U, de la Chapelle A, Schlessinger D (1996) Nat Genet 13:409
207. Chen Y, Molloy SS, Thomas L, Gambee J, Bächinger HP, Ferguson B, Zonana J, Thomas G, Morris NP (2001) Proc Natl Acad Sci USA 98:7218
208. Oh SP, Kamagata Y, Muragaki Y, Timmons S, Ooshima A, Olsen BR (1994) Proc Natl Acad Sci USA 91:4229
209. Halfter W, Dong S, Schurer B, Cole GJ (1998) J Biol Chem 273:25404
210. Li D, Clark CC, Myers JC (2000) J Biol Chem 275:22339
211. Myers JC, Dion AS, Abraham V, Amenta PS (1996) Cell Tissue Res 286:493
212. Muragaki Y, Timmons S, Griffith CM, Oh SP, Fadel B, Quertermous T, Olsen BR (1995) Proc Natl Acad Sci USA 92:8763
213. Saarela J, Rehn M, Oikarinen A, Autio-Harmainen H, Pihlajaniemi T (1998) Am J Pathol 153:611
214. Eklund L, Piuhola J, Komulainen J, Sormunen R, Ongvarrasopone C, Fässler R, Muona A, Ilves M, Ruskoaho H, Takala TES, Pihlajaniemi T (2001) Proc Natl Acad Sci USA 98:1194
215. Fukai N, Eklund L, Marneros AG, Oh SP, Keene DR, Tmarkin L, Niemelä M, Ilves M, Li E, Pihlajaniemi T, Olsen BR (2002) EMBO J 21:1535
216. Marneros AG, Olsen BR (2003) Invest Ophthalmol Vis Sci 44:2367
217. Sertie AL, Sossi V, Camargo AA, Zatz M, Brahe C, Passos-Bueno MR (20000 Hum Mol Genet 9:2051
218. O'Reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, Flynn E, Birkhead JR, Olsen BR, Folkman J (1997) Cell 88:277
219. Sasaki T, Fukai N, Mann K, Göhring W, Olsen BR, Timpl R (1998) EMBO J 17:4249
220. Wen W, Moses MA, Wiedershain D, Arbiser JK, Folkman J (1999) Cancer Res 59:6052
221. Felbor U, Dreier L, Bryant RAR, Ploegh HL, Olsen BR, Mothes W (2000) EMBO J 19:1187
222. Ferreras M, Felbor U, Lenhard T, Olsen BR, Delaissé J-M (2000) FEBS Lett 486:247
223. John H, Preissner KT, Forssmann W-G, Ständker L (1999) Biochemistry 38:10217
224. Sasaki T, Larsson H, Tisi D, Claesson-Welsh L, Hohenester E, Timpl R (2000) J Mol Biol 301:1179

225. Ramachandran R, Dhanabal M, Volk R, Waterman MJF, Segal M, Lu H, Knebelmann B, Sukhatme VP (1999) Biochem Biophy Res Commun 255:735
226. (a) Huggins LG, Lennarz WJ (2001) Dev Growth Differ 43:415; (b) Dankwrdt SM, Billedeau RJ, Lawley LK, Abbot SC, Martin RL, Chan CS, Van Wart HE, Walker KA (2000) Bioorg Med Chem Lett 10:2513; (c) Dankwardt SM, Martin RL, Chan CS, Van Wart HE, Walker KA, Delaet NG, Robinson LA (2001) Bioorg Med Chem Lett 11:2085; (d) Dankwardt SM, Abbot SCV, Broka CA, Martin RL, Chan CS, Springman EB, Van Wart HE, Walker KA (2002) Bioorg Med Chem Lett 12:1233; (e) Delaet NG, Robinson LA, Wilson DM, Sullivan RW, Bradley EK, Dankwardt SM, Martin RL, Van Wart HE, Walker KA (2003) Bioorg Med Chem Lett 13:2101; (f) Robinson LA, Wilson DM, Delaet NG, Bradley EK, Dankwardt SM, Campbell JA, Martin RL, Van Wart HE, Walker KA, Sullivan RW (2003) Bioorg Med Chem Lett 13:2381

Top Curr Chem (2005) 247: 185–205
DOI 10.1007/b103823
© Springer-Verlag Berlin Heidelberg 2005

Collagen Suprastructures

David E. Birk[1] (✉) · Peter Bruckner[2]

[1] Department of Pathology, Anatomy & Cell Biology, Thomas Jefferson University, 1020 Locust Street, JAH 543, Philadelphia PA19107, USA
david.birk@jefferson.edu

[2] Institute for Physiological Chemistry and Pathobiochemistry, University Hospital of Münster, Waldeyerstrasse 15, 48149 Münster, Germany
peter.bruckner@uni-muenster.de

1 Introduction 186

2 Fibrillar Collagens 187
2.1 Polymerization of Fibrillar Collagens 187
2.2 Influence of Post-Translational Processing of Collagens on Fibril Structure . 189

3 Fibril-Associated Collagens with Interrupted Triple Helices (FACIT) 189

4 Network-Forming Collagens 190
4.1 Collagen IV Networks 190
4.2 Collagen VI Networks 192
4.3 Collagen VIII and X Networks 194

5 Assembly of Collagen Suprastructures in Tissues 196
5.1 Cartilage Fibril Formation 196
5.2 Corneal Fibril Formation 198
5.3 Tendon Fibril Formation 201

6 Concluding Remarks 202

References 203

Abstract Collagens are a family of proteins with considerable molecular diversity. The genetic diversity is increased by assembly of multiple isoforms, alternative splicing and post-translational modifications. All collagen molecules share common features, including having at least one domain composed of three polypeptide chains organized as a triple helix. However, these molecules assemble to form a variety of supramolecular structures, e.g., fibrils, filaments, or networks, responsible for the characteristics of specific extracellular matrices. These supramolecular structures are alloys or macromolecular composites, containing different collagen types as well as other matrix macromolecules. This heteropolymeric composition provides an additional level of complexity. Current concepts of collagen suprastructures, their assembly and function within extracellular matrices will be the focus of this review.

Keywords Collagen supramolecular structures · Extracellular matrix · Fibrillar collagens · Network-forming collagens · FACIT collagens

1 Introduction

At least 27 known types of collagens constitute a family of glycoproteins that occur in extracellular matrices of vertebrates and invertebrates with different suprastructural organizations (Table 1). All collagens are trimers, each composed of one, two or three distinct gene products. The molecular diversity of collagen polypeptides is enhanced further by alternative promoters and mRNA-splicing as well as by the possibility to combine collagenous polypeptides in alternative manners that, for some collagen types, results in the formation of isoforms. Finally, collagen molecules are subject to a variable extent of post-translational modification and/or proteolytic processing, modulating even further their molecular architectures and surface properties (for review, see chapter Ricard-Blum).

Collagen molecules have in common at least one domain comprising three polypeptide chains folded into collagen-like triple helices. This hallmark feature renders collagen molecules well adapted to assemble into highly multimeric suprastructural aggregates, thereby converting protomeric, mainly unfunctional molecules into their operative state. In this respect, it is particularly intriguing that one and the same collagen type can predominate, sometimes overwhelmingly, in tissue aggregates that otherwise are strikingly diverse. This is well exemplified by collagen I, the major protein of banded fibrils with vastly dissimilar tissue-specific organizations. The most likely explanation for this paradox is the fact that most, if not all, collagen-containing suprastructures have complex macromolecular compositions that not only include other collagen types, but also non-collagenous components. These additional macromolecules may be substantial or occur in minute quantities. Invariably, however, they decisively influence tissue-specific architectures and functions. This review will outline our current understanding of the compositional requirements of supramolecular assemblies containing collagens to yield fibrils, microfibrils, and/or networks. We shall dwell less on the molecular collagen structures, a subject covered elsewhere in this volume as well as in recent excellent reviews

Table 1 Collagen suprastructures

Suprastructure	Collagen types
Fibril	I, II, III, V, XI, XXIV, XXVII
Fibril-associated (FACIT[a])	IX, XII, XIV, XVI, XIX, XX, XXI, XXII
Network	IV, VI, VIII, X
Anchoring fibrils	VII
Transmembrane collagens	XIII, XVII, XXIII, XXV
Multiplexin	XV, XVIII

[a] Fibril-associated collagens with interrupted triple helices.

[1–3]. Tissue suprastructures, like metal alloys, are uniform materials comprising several molecular species and with properties distinct from those of unimolecular aggregates. This is well exemplified by cartilage fibrils (see below). However, collagen-containing suprastructures also can be composites of more than one alloy or unimolecular assemblies. As also detailed below, basement membranes represent good examples for this mode of aggregation. In the following, we shall discuss current concepts of the supramolecular structure, assembly and function within extracellular matrices.

2 Fibrillar Collagens

2.1 Polymerization of Fibrillar Collagens

The fibril-forming collagens, i.e., I, III, V, XI, XXIV and XXVII are rod-like collagens with a large triple helical domain (ca. 300 nm) containing 990 to 1020 amino acids per polypeptide chain. It is now clear that these collagens co-assemble into banded fibrils in tissues including bone, dentin, tendons, cartilage, dermis, sclera, cornea, and the interstitial connective tissues in and around many organs. They are arranged in longitudinally staggered arrays of molecules of a length that is a non-integer multiple of the stagger between next neighbors. Thus, a gap occurs sequentially between neighboring molecules giving rise to a gap-overlap structure in all collagen fibrils with a D-periodic banding as schematically represented in Fig. 1. In addition, the longitudinal axis of the fibrillar collagen molecules is not parallel to that of the fibrils. It still is unclear whether the molecules are supercoiled around each other or whether they are crimped within the fibrils [4]. It is possible that, depending on the type and the origin of fibrils, both forms of organization exist. The relative abundance of the collagen types in fibrils is tissue- or tissue domain-specific and, in some cases, may be very small. For example, collagen III has been reliably identified in certain bone fibrils by immunoelectron microscopy [5], but was not detectable among the collagen proteins extracted from that tissue. Either, collagen III is too scarce to allow visualization by protein chemical techniques or, more likely, is insoluble due to the highly abundant covalent cross-linking. Nevertheless, collagen III is likely to affect fibrillar assembly and structure because its triple helical domain is longer than that of type I, the major bone collagen. The triple helical domain of collagen III is composed of 340 amino acid triplets Gly-Xaa-Yaa per chain whereas that of collagen I only has 338. In addition to type III, a bone-specific variant of collagen V/XI has been identified in bone [6]. Likewise, the molecular composition of skin fibrils is complex even in terms of fibrillar collagens. Again, the major collagen is type I, that is co-polymerized with substantial amounts of collagen III in a developmentally regulated and/or tissue domain-dependent manner. In addition, lesser amounts of other fibrillar

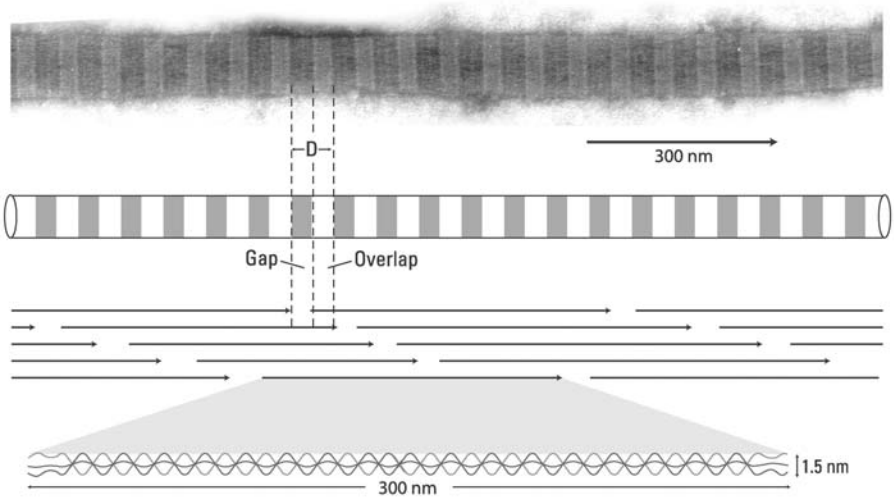

Fig. 1 Structure of a generic collagen fibril. A D-periodic collagen fibril from tendon is presented at the *top of the panel*. The negative stained fibril has a characteristic alternating light/dark pattern representing the gap (*dark*) and overlap (*light*) regions of the fibril. The diagram represents the staggered pattern of collagen molecules giving rise to this D-periodic repeat. The collagen molecules (*arrows*) are staggered N to C. The fibrillar collagen molecule is approximately 300 nm (4.4 D) in length and 1.5 nm in diameter

collagens, mainly type V, also are present in skin fibrils. Again, it is likely that distinct fibrils arise from these mixtures and that the distinctions depend on the molar fractions of the composite fibrillar collagens.

One supramolecular characteristic of banded collagen fibrils is the length of the stagger between adjacent fibrillar collagen molecules (D-period) that is tissue-specific (e.g., 67 nm in rat tail tendon and 64 nm in human dermis) [7]. Another variable between tissues is the tilt of the long axis of collagen molecules with respect to the longitudinal axis of the fibril. Variations of this kind and magnitude necessitate substantial differences in molecular organization, including longitudinal D-periodic packing as well as intermolecular distances. It has recently been shown, that very small quantities of collagen XI (less than 1 part in 1000) can profoundly alter the fibrillar organization of collagen I. In addition, fibril formation by pure collagen I in vitro is remarkably slow and incompatible with the requirements of fibrillogenesis in situ because nucleation in the absence of collagen V/XI represents a formidable kinetic barrier against the assembly process. Thus, considerable lag periods of aggregation ensue. However, the quantitatively minor collagen XI, together with collagen I, can form a uniform core of the fibril that efficiently nucleates further lateral growth by accretion of collagen I. Thus, collagen I/XI-fibrils are composites containing alloyed cores and sheaths of pure collagen I. Their final morphology is variable and depends on the collagenous composition [8]. Extrapolating this concept to banded fibrils in general, it has now become clear that the mixtures of fibrillar

collagens, i.e., types I, II, III, V/XI, XXIV and XXVII, act as key regulators of the collagen organization in fibrils, resulting in different modes of supercoiling or crimping and, hence, the length of the D-periodic banding.

2.2
Influence of Post-Translational Processing of Collagens on Fibril Structure

The covalent modifications occurring after polypeptide synthesis are particularly prominent in collagens and are likely to have a profound impact on the assembly of fibrillar collagens [9–11], as well as FACITs. The cylindrical circumference of triple helical domains of collagens is affected by the varying extent of hydroxylation of lysyl residues and subsequent galactosylation and glucosyl-galactosylation of hydroxylysyl residues. Thus, intermolecular center-to-center distances correlate with the extent of glycosylation, especially if the post-translational modifications affect polypeptide parts eventually situated in overlap regions of the collagen packing. The extent of glycosylation of hydroxylysine can be manipulated by temperature [9], activity levels of modifying enzymes in the relevant subcellular compartments [12], or, most notably, by disease-causing mutations [2]. Such mutations can substantially reduce the rates of triple helix-formation in fibrillar procollagens in the rough endoplasmic reticulum and cause overmodification because the hydroxylating and glycosylating enzymes reside in that compartment and only recognize unfolded collagen-like polypeptides as their substrate. Collagen molecules that are mutated and overmodified in this manner co-polymerize with normal molecules, thereby compromising normal fibrillar organization. This molecular relationship of genotypes and phenotypes has thus been designated as "protein suicide". However, differences in the extent of glycosylation also can be a mode of physiological regulation of fibrillar organization. For example, collagen I is known to be glycosylated in regions residing in the overlap domains of banded fibrils if the protein is a product of corneal, but not tendon fibroblasts.

An attractive possibility of regulation of fibril morphology is that the availability of protomers by enzymatic cleavage may control fibril growth and shape [13]. However, it is not easy to see how a kinetic control can determine long-term stability of fibrillar assemblies. The heterotypic fibril assembly model also involving non-collagenous components, such as the small leucine-rich proteoglycan decorin, has the advantage of introducing the possibility of thermodynamic shape control [14]. Nevertheless, the two concepts may act in concert.

3
Fibril-Associated Collagens with Interrupted Triple Helices (FACIT)

The supramolecular complexity of D-banded fibrils is augmented further by the incorporation of fibril-associated collagens with interrupted triple helices (FACIT). This collagen subfamily comprises the types IX, XII, XIV, XVI, XIX,

XX, XXI, XXII, and XXVI, and its members are structurally very diverse. They only have in common a FACIT-domain, i.e., a relatively short, carboxy-terminal triple helical stretch flanked by a cysteine-containing motif, GXCXXXC, at the junction of the triple helix and the carboxy-terminal non-helical region. The cysteines are essential for the covalent bonding of the three constituent polypeptides. It has been proposed that the FACIT-domain of collagen IX may be incorporated into the gap between consecutive fibrillar collagen molecules in cartilage fibrils [15]. Although this has not formally been proven it is an attractive possibility that other FACITs are integrated into corresponding fibrils in other tissues and, therefore, provide an opportunity for tissue-specific modulation of the fibril surfaces by projection of the non-FACIT-domains into the perifibrillar space. The selective expression of the FACITs would thus provide an elegant molecular mechanism for molding of fibril surfaces. Indeed, it has been shown that the amino-terminal regions of collagens IX, or XII and XIV protrude from the fibril surfaces in cartilage or skin and tendon, respectively [5, 16]. The biomechanical diversity of banded fibrils may thus be a direct consequence of distinctions of surface properties afforded by FACITs.

At least in the case of the prototypic FACIT, collagen IX, triple helical domains other than the carboxyterminal FACIT-domain also are incorporated into the fibril body, possibly by an antiparallel alignment with the fibrillar collagens of cartilage (see below). In doing so, FACITs become part of the fibril bodies, very much like the fibrillar collagens, thereby modulating further and/or stabilizing the molecular organization within fibrils. However, the suprastructural association with fibrillar collagens is not a generalized feature of all FACITs. For example, collagen XVI can be an optional constituent of banded fibrils in cartilage (see below). In skin, however, the protein never is incorporated into banded fibrils, but rather is a component of fibrillin-containing microfibrils. Recently, we showed that in tissue junctions collagen XXII also coexists with fibrillin-containing suprastructures rather than banded fibrils [17]. Thus, FACITs are not always are associated with banded fibrils and, where they are, they can be integral parts and important organizers of the overall fibril structure rather than optional additions to preexisting aggregates.

4 Network-Forming Collagens

4.1 Collagen IV Networks

Basement membranes are composites consisting of several independent, but intertwined supramolecular networks. One of these networks contains collagen IV that comes in several isoforms, whereas the others harbor as their major molecular components laminin, also existing in several isoforms, or perlecan, respectively. Further macromolecules, including nidogen/entactin, have been

shown to mediate molecular contacts, thereby stabilizing the compound macromolecular networks in basement membranes. Distinct basement membranes occur in different anatomical locations or may be sequentially formed during development. This also is reflected by the subtypes of collagen IV that are differentially expressed under distinct circumstances. There are six collagen IV-encoding genes, i.e. COL4A1 through COL4A6, giving rise to the corresponding polypeptides alpha1(IV) through alpha6(IV). There is only a limited set of heterotrimeric molecules folded into triple helical molecules with the stoichiometries [alpha 1(IV)]$_2$ alpha 2(IV), alpha 3(IV) alpha 4(IV) alpha 5(IV), and [alpha 5(IV)]$_2$ alpha 6(IV), respectively. Other chain combinations apparently do not exist. Collagen IV molecules aggregate into networks by interactions between their N-terminal, triple helical domains, called 7S-domains, organizing 4 similar collagen IV molecules in an antiparallel fashion. From such knots formed by 7S-domains the long collagen IV-triple helices stretch out into all spatial directions. This flexibility is afforded by interruptions in the typical (Gly-X-Y)$_n$-sequences that, unlike in fibrillar collagens, occur abundantly in collagen IV molecules, including a site between the small 7S- and the long triple helices. Such interruptions create a point of flexibility in an otherwise stiffly rod-like molecule. At their C-terminus, non-collagenous NC1-domains interact head-to-head, creating in conjunction with the 7S-interactions large supramacromolecular aggregates resembling chicken-wire-like interwoven networks. In the interaction between collagen IV – molecules involving NC1-domains, heterotypic arrangements are possible. [alpha 1(IV)]$_2$ alpha 2(IV)-trimers can interact with [alpha 5(IV)]$_2$ alpha 6(IV) trimers by interactions between alpha1- and alpha5-, as well as alpha 2- and alpha 6-NC1 domains. By contrast, alpha 3 (IV) alpha 4 (IV) alpha 6 (IV) interact through their NC1-domains to yield pairs of 2 alpha 4 (IV)-NC1-domains or alpha 3 (IV)-alpha 5 (IV)-NC1-heterodimers. Thus, heterotypic networks arise with distinct supramolecular structures and propensities to undergo further aggregation reactions. Such chicken-wire-like networks undergo further supramolecular assembly by laterally aggregating into extended polygonal networks with widely variable mesh-sizes.

Suprastructures formed by various combinations of collagen IV-isoforms can undergo further molecular interactions with basement membrane-associated macromolecules. These include collagen VII-containing anchoring fibrils in the dermo-epidermal junction zone or collagen XVIII/endostatin-containing filamentous suprastructures associated with basement membranes underlying the retinal pigment epithelium or in Bruch's-membrane in the eye as well as in blood vascular endothelial basement membranes [18]. Mutations in collagens VII and XVIII destabilizing the interactions of these molecules with basement membrane components lead to dermal blistering diseases (dystrophic epidermolysis bullosa) or to Knobloch-syndrome, a rare disease characterized by severe ocular alterations, including vitreoretinal degeneration associated with retinal detachment and occipital scalp defect [19]. In each of these cases, the stability of basement membrane zones are destabilized by the mutations.

4.2 Collagen VI Networks

Collagen VI is an ubiquitous component of connective tissues. It is found as an extensive filamentous network with collagen fibrils and is often enriched in pericellular regions. It is assembled into several different tissue forms, including beaded microfibrils, hexagonal networks and broad banded structures [20, 21]. Collagen VI interacts with a spectrum of extracellular molecules including: collagen types I, II, IV, XIV, microfibril-associated glycoprotein (MAGP-1), pearlecan, decorin, biglycan, hyaluronan, heparin and fibronectin as well as integrins and the cell-surface proteoglycan NG2. Based on the tissue-localization and large number of potential interactions, collagen VI has been proposed to integrate different components of the extracellular matrix, including cells [22]. In addition, collagen VI may influence cell migration, differentiation and apoptosis/proliferation. This indicates a role(s) in the development of tissue-specific extracellular matrices, repair processes and in the maintenance of tissue homeostasis.

Collagen VI is a heterotrimer composed of alpha1(VI), alpha2(VI) and alpha3(VI) chains [22, 23]. The type VI monomer has a 105 nm triple helical domain with flanking N- and C-terminal globular domains. The N-terminal domain is almost exclusively from the alpha3(VI) chain and has approximately twice the molecular mass of the C-terminal domain. The N- and C-terminal domains are composed of varying numbers of von Willebrand type A repeats. The alpha3(VI) N-terminal domain is larger than the comparable domains in the alpha1(VI) and alpha2(VI) chains, with a maximum of ten type A repeats vs one. The alpha3(VI) C-terminal domain has five type A repeats vs two for the alpha1(VI) and alpha2(VI) chains. The terminal type A repeat of the alpha3(VI) chain can be processed extracellularly. Structural heterogeneity is introduced by alternative splicing of domains, primarily of the alpha3(VI) N-terminal domain. This domain is expressed in several different forms giving rise to structural and functional heterogeneity.

A distinct property of collagen VI is that assembly of the supramolecular forms begins in the lumen of intracellular compartments (Fig. 2). The initial assembly step is dimer formation via the lateral, anti-parallel association of two monomers. The monomers are staggered by 30 nm with the C-terminal domains interacting with the helical domains. This gives rise to an overlapped, central 75 nm helical domain flanked by a non-overlapped region with the N- and C-globular domains, each about 30 nm. These interactions are stabilized by disulfide bonds near the ends of the overlapped region [24]. The overlapped helices of the two monomers form a supercoil in the central region [25]. In the next step, tetramers form when two dimers align with the ends in register. The second (C2) C-terminal type A repeat in the alpha2(VI) chain is critical for dimer and subsequent tetramer formation. This involves an interaction between the C2 region of one alpha2(VI) chain and the helical region of the anti-parallel alpha2(VI) chain [26]. Tetramers are secreted and are the building blocks that assemble extracellularly into the tissue forms of type VI collagen.

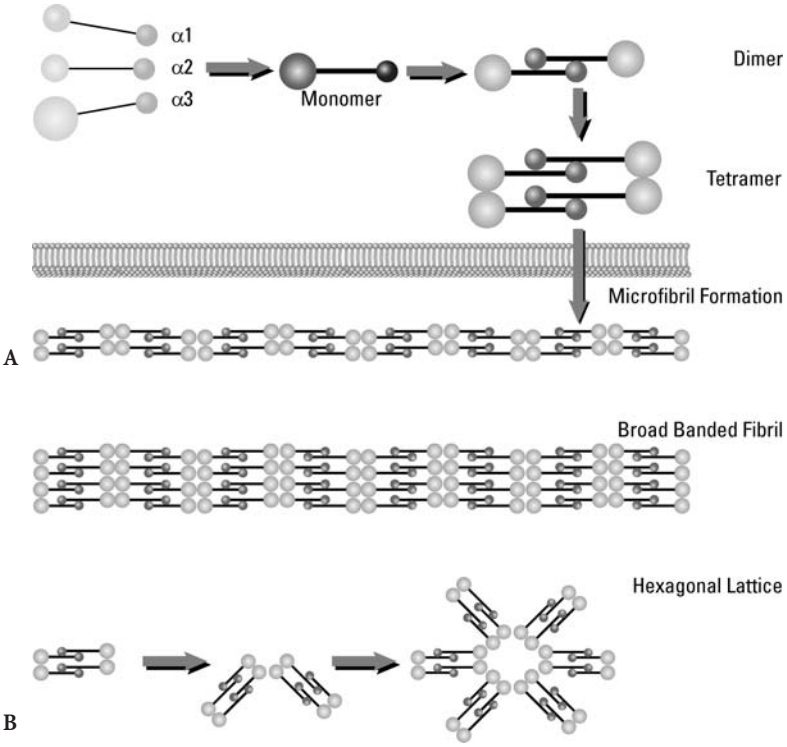

Fig. 2A, B Assembly of collagen VI suprastructures: **A** collagen VI forms tetramers intracellularly. The collagen VI monomer assembles from 3 alpha chains and has a C-terminal globular domain, a central triple helical domain and an N-terminal globular domain. The monomers assemble N-C to form dimers. Tetramers are assembled from two dimers aligned in register. The tetramers are secreted and form the building blocks of three different collagen VI suprastructures; **B** beaded filaments, broad banded fibrils and hexagonal lattices form via end-to-end interactions of tetramers and varying degrees of lateral association. (Diagrams modified from [28, 30])

In the extracellular environment, collagen VI tetramers assemble to form the suprastructures found in tissues. Tetramers associate end-to-end forming beaded filaments. This is a non-covalent interaction that is presumably mediated through Type A domain interactions. This gives rise to thin, beaded filaments (3–10 nm) with a periodicity of approximately 100 nm. These beaded filaments laterally associate, forming beaded microfibrils [20, 27, 28]. In addition to beaded microfibrils, other type VI collagen-containing supramolecular structures are found in the extracellular matrix including hexagonal lattices; and broad banded fibrils with a 100 nm periodicity. The hexagonal lattices are formed via end-to-end interactions of tetramers in a non-linear fashion. While the broad banded fibrils probably represent continued lateral growth of beaded microfibrils and/or lateral association of preformed beaded microfibrils (Fig. 2).

Supramolecular aggregates of collagen VI are composite structures. As discussed above, collagen VI interacts with a large number of extracellular matrix molecules. Comparable to collagen fibrils, supramolecular aggregates containing collagen VI are composite structures with other integrated molecules modulating the functional properties of the type VI collagen-containing suprastructure. For example, biglycan and decorin bind to the same triple helical site near the N-terminus [29]. This interaction is mediated by the core protein and the presence of the glycosaminoglycan chain(s) had no effect on binding. Biglycan interactions with the tetramer induced formation of hexagonal lattices rather than beaded microfibrils. This was dependent on the presence of the glycosaminoglycan chains. In contrast, decorin, which binds to the same site, was less effective in inducing hexagonal lattice formation [30]. Analogous to fibril formation, the interaction of small leucine-rich proteoglycans with collagen VI influences the structure of the tissue aggregate and therefore its function. In addition, this regulation involves two closely related, class I, leucine-rich proteoglycans that compete for the same binding site. In cartilage, the expression of biglycan is enriched in the pericellular environment while decorin is enriched in the territorial matrix. This provides a mechanism to assemble different suprastructures in adjacent regions or tissues with different functions. Coordinate changes in expression patterns could influence matrix assembly during development and repair. In addition, abnormal changes in expression would alter the tissue-specific suprastructure and may lead to tissue pathology. In addition, collagen VI-associated decorin or biglycan can form complexes with matrilin-1. In the cartilage matrix, these complexes mediate interactions between the collagen VI network, the fibrillar network and the aggrecan network [31]. This illustrates another mechanism whereby the composite structure of collagen suprastructures contributes to define the specific structure/function associated with different tissues.

4.3 Collagen VIII and X Networks

Collagens VIII and X assemble to form hexagonal networks in tissues. These collagens are closely related, with comparable gene and protein structures [1, 22, 32]. Collagen VIII is a major component of blood vessels, located subendothelially and Descemet's membrane, separating the corneal endothelium from stroma [33, 34]. Descemet's membrane is composed of layers of hexagonal lattices [35]. These lattices are suprastructures containing collagen VIII [34]. Collagen X has a very restricted distribution, found only in hypertrophic cartilage. The supramolecular form is a hexagonal lattice containing collagen X similar to that formed by collagen VIII [36]. Collagen VIII is a homo or hetero-trimer of alpha1(VIII) and alpha2(VIII) chains while collagen X is composed of a single alpha1(X) chain. There is evidence that both collagen VIII homotrimers and the $[\text{alpha1(VIII)}]_2$alpha2(VIII) heterotrimer exist in tissues [37, 38].

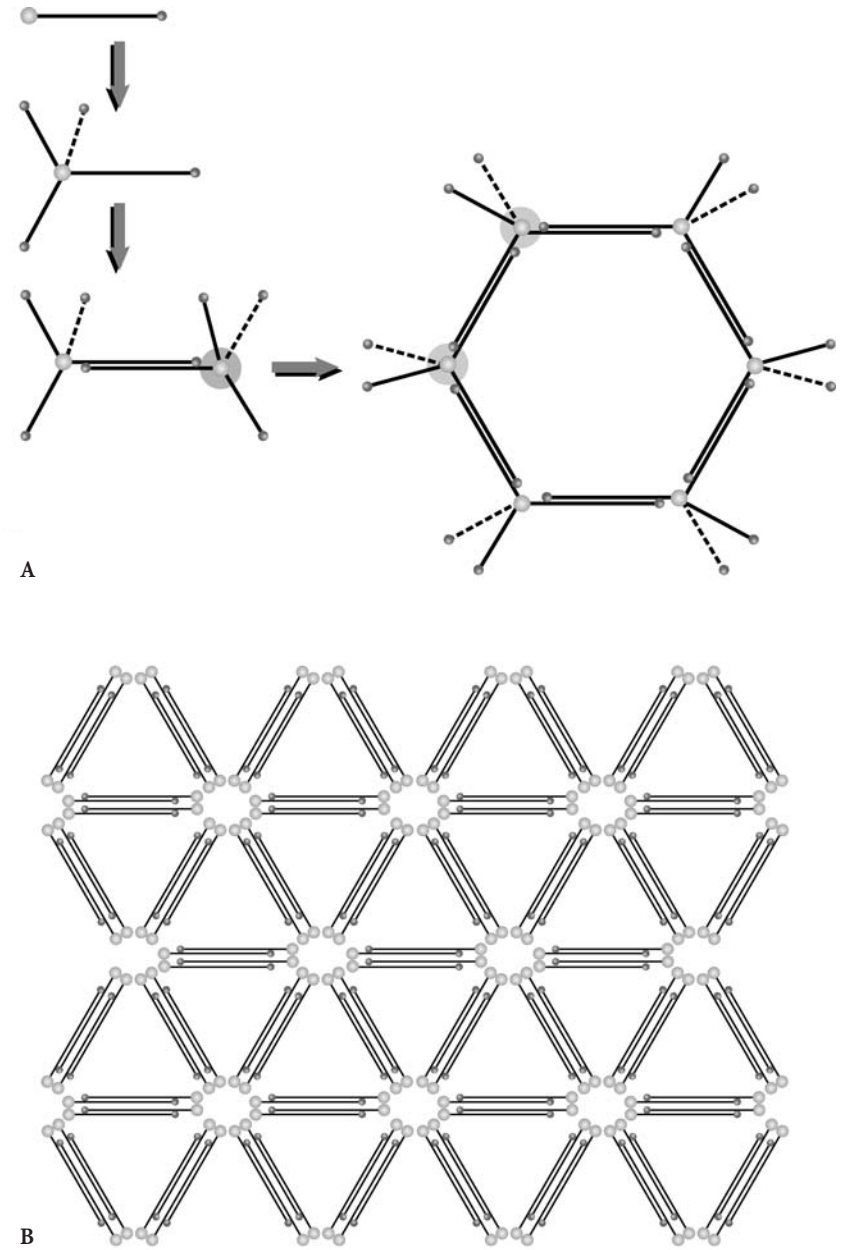

Fig. 3A, B Assembly of collagen VIII hexagonal lattices: **A** the C-terminal, non-collagenous domains of four collagen VIII molecules interact to form tetrahedrons. Tetrahedrons assemble further to form hexagonal lattices; **B** a planar hexagonal lattice is diagrammed. Continued assembly, involving interactions of the amino terminal non-collagenous domains or anti-parallel interactions involving both helical and terminal domains would generate a layered hexagonal lattice. (Model modified from [39])

Recently, collagen VIII lattices have been assembled in vitro [39]. These in vitro lattices are comparable to those seen in tissues and a model for assembly of the collagen VIII suprastructure was proposed (Fig. 3). Collagen VIII molecules form a tetrahedron through the interaction of four molecules. The interaction is proposed to involve the carboxy non-collagenous domains that have a conserved hydrophobic patch [40]. This structure is proposed as the building block that assembles into three-dimensional hexagonal lattices. The assembly of a layered hexagonal lattice could involve interaction of the amino terminal non-collagenous domains or anti-parallel interactions involving both helical and terminal domains. The anti-parallel interactions are consistent with the thicker inter-nodal struts observed and it is predicted that this is the primary mechanism for formation of hexagonal lattices both in vitro and in tissues.

5
Assembly of Collagen Suprastructures in Tissues

5.1
Cartilage Fibril Formation

D-periodically banded (D=64 nm) fibrils in cartilage [41] constitute the major fraction of the dry weight of the matrix in this tissue and are the main tensile element containing the swelling pressure generated by osmotic binding of water to the highly polyanionic glycosaminoglycan chains of the extrafibrillar matrix. Two major populations of fibrils exist in cartilage. There are thin fibrils with a uniform diameter of about 20 nm that occur throughout the extracellular matrix of all hyaline cartilages. They are particularly enriched in the territorial matrix where they have a preferential orientation parallel to the surface of the chondrocytes. Thus, the fibrils embed and separate individual chondrocytes by forming basket-like structures [42]. The second population almost exclusively occurs in specialized matrix compartments, termed interterritorial regions, and are more remote from the chondrocytes. In growth plates and in articular cartilage, their preferential orientation is parallel to the long axis of the bones and the direction of forces generated by load bearing. It is unclear by what mechanism the wider fibrils are formed, but it is probable that the thin territorial fibrils correspond to the prototypic cartilage fibrils that, after appropriate processing, can undergo fusion to form the larger interterritorial fibrils. The other lateral growth mechanism, i.e., the direct apposition of collagens and other macromolecules to pre-existing thin fibrils, is less likely to operate since the cells producing the macromolecular fibril constituents are separated by large distances from the thick and well banded fibrils of the interterritorial matrix. This would necessitate extensive diffusion of fibril macromolecules through the dense network of cartilage matrix.

Cartilage fibrils exquisitely illustrate the concept of matrix aggregates as macromolecular composites/alloy. Their quantitatively major component is the

fibril-forming collagen II, a protein that typically, but not exclusively, occurs in all types of cartilage. The designation of collagen II as a "fibril-forming" collagen is appropriate in the sense that the protein is an essential fibril component in cartilage. However, the protein, by itself, is incapable of forming fibrils of the extensive lengths required in the tissues. Instead, collagen II forms tactoidal structures with a strong D-banding pattern and with very limited lengths when the pure protein is subjected to aggregation in vitro. The tactoids essentially consist of two tapering ends joined together back-to-back. In addition, they are only formed at very high initial concentrations of monomeric collagen II and lack any obvious lateral growth control. It is most likely that, for these reasons, collagen II never occurs as a single protein in cartilage fibrils in situ. Rather, the protein always occurs in macromolecular composites. In the case of the prototypic, territorial fibrils, the collagenous components include collagens II, IX, and XI [43] or, more rarely, collagens II, XI, and XVI [44]. When subjected to fibrillogenesis in vitro, mixtures of collagens II and XI alone exhibit an outstanding capability of forming thin and uniform fibrils with a diameter of about 20 nm and closely resembling cartilage fibrils in the electron microscope. Interestingly, this tight diameter control is observed only when collagens II and XI are present at molar fractions $f_{II/XI}$=[collagen II]/[collagen XI]≤8, a proportion strikingly similar to that occurring in authentic prototypic fibrils. Thus, the mode of lateral packing in prototypic fibrils is jointly dictated by collagens II and XI forming a uniform mixture and, for this reason, can be likened to a metal alloy. However, the fibrils formed in vitro by collagens II and XI, alone, appear to be somewhat less tightly packed than prototypic fibrils. In addition, they are less stable in that the two collagens lose their aggregating capacity upon prolonged standing without demonstrable proteolytic alteration. Moreover, this loss of competence for fibrillogenesis is readily rescued by the addition to the reconstitution mixtures of collagen IX. Thus, collagen IX is essential for the overall formation of cartilage fibrils with long-term stability rather than to be a "decorative" addition to aggregates preformed by the other two collagen types. Therefore, collagen IX is to be considered as the third collagenous component of the macromolecular alloy in cartilage fibrils and can be important during assembly or after assembly in a tissue-specific manner [45]. Such composite fibrils comprise a heterotypic fibril body encompassing parts of all three collagen types from which collagen IX can project to the exterior its aminoterminal globular NC4-domain with the collagenous region Col3 serving as a spacer. The protein also may interconnect individual cartilage fibrils in the tissue [15, 46].

The macromolecular components of cartilage fibrils are unlikely to be restricted to collagens. When subjected to electrophoresis in polyacrylamide gels after extensive denaturing treatment (SDS, reducing agents, boiling for extended periods of time), prototypic fibrils generate diffuse electrophoretic patterns resembling those of proteoglycans. These smears arise from abundant and tightly bound fibril constituents that, unlike the collagens, are susceptible to degradation by certain exogenously added proteinases [16]. The identity of

these non-collagenous fibril components still is unknown. Conceivably, they include glycoproteins or proteoglycans such as the members of the family of proteins with leucine-rich repeat motifs. In fact, decorin is known to occur preferentially in the interterritorial zones of cartilage matrix that is rich in the largest banded fibrils lacking collagen IX as visualized by immunoelectron microscopy [47]. This led to the hypothesis that interterritorial banded fibrils arise from small prototypic collagen IX-containing fibrils after proteolytic elimination of collagen IX or, at least, its aminoterminal Col3- and NC4-domains studding the fibril surface. After accommodation of such processed prototypic fibrils into the D-periodic stagger, fusion is thought to occur accompanied by a polyanionic conditioning of the fibril surface by incorporation of the proteoglycan decorin. This mechanism of lateral growth would eliminate the necessity of diffusion of procollagens and procollagen-processing proteinases over large expanses of dense fibrillar cartilage networks. However, it would also necessitate the selective presence of fibril-processing proteinases at the boundaries of territorial and interterritorial zones of cartilage matrix. A challenging task of the future will be the identification of the hypothetical proteinases and the elucidation of their regulated action.

5.2 Corneal Fibril Formation

The mature corneal stroma is composed of a single, homogeneous population of small diameter collagen fibrils organized as orthogonal layers. This stroma provides for the mechanical stability of the anterior eye and for corneal transparency. Both of these properties are dependent on the supramolecular organization of collagen into the composite fibrils characteristic of this extracellular matrix. The corneal fibrils are alloys assembled from collagens I and V. These collagen I/V fibrils interact with FACIT collagens, types XII and/or XIV depending on developmental stage. In addition, the small leucine rich proteoglycans, decorin, lumican, keratocan, and osteoglycine interact with the fibril surface (Fig. 4). These heteropolymeric fibrils provide a tissue-specific composite fibril that is responsible for the unique properties of the cornea. The formation of this supramolecular aggregate is a multistep assembly process involving: fibril initiation, initial assembly of a fibril intermediate, followed by linear fibril growth with a lack of lateral fibril growth. Fibril initiation and initial assembly of a fibril intermediate is dependent on the collagen-collagen interactions that produce the alloy, while the later fibril growth steps are dependent on fibril-associated macromolecules.

Fibril assembly involves collagen-collagen interactions. The corneal keratocytes synthesize two fibril forming collagens. Collagen I is the quantitatively major collagen making up 80–90% of the total while the $[\text{alpha1(V)}]_2$ alpha2(V) isoform of collagen V is the quantitatively minor fibril-forming collagen [48, 49]. These two collagens co-assemble to form a heterotypic fibril [50, 51]. This co-assembly regulates the formation of the initial fibril intermediate. In addition,

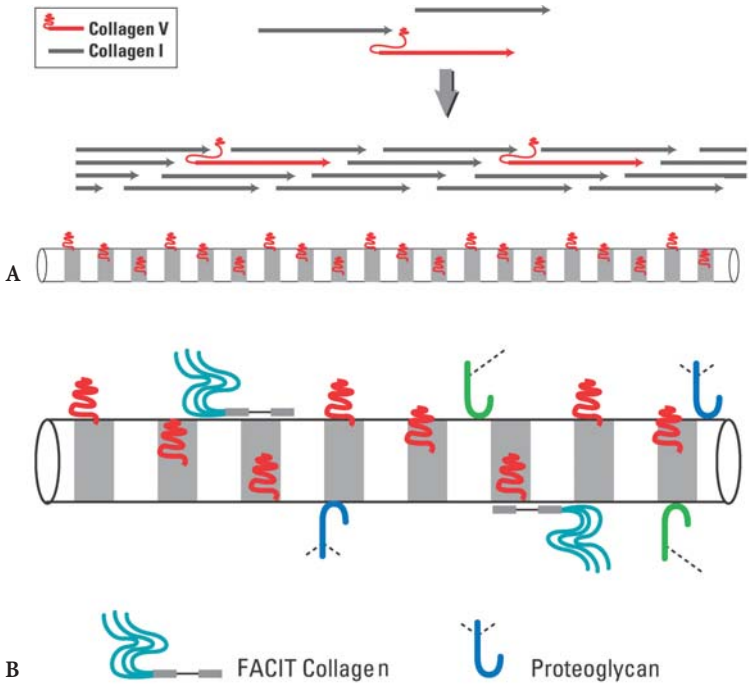

Fig. 4A, B Corneal fibril: **A** corneal fibrils are heterotypic, co-assembled from collagens I and V. Collagen V is a quantitatively minor component of most collagen I-containing fibrils. It has a retained N-terminal, non-collagenous domain that must be in/on the gap region/fibril surface. The heterotypic interaction is involved in efficient initiation of fibril assembly; **B** the heterotypic alloy forms the core of a composite fibril with fibril-associated leucine-rich repeat proteoglycans and FACIT collagens bound to the surface. While the heterotypic composition is relatively constant, the fibril-associated macromolecules are more dynamic, changing temporally during development or repair and spatially in different tissues/tissue domains

corneal type I collagen differs from that found in other tissues in its elevated level of glycosylation [52]. This post-translational modification also affects regions of collagen I that are incorporated into the overlap zones of fibrils. The increase in diameter of individual collagen molecules introduced by the galactose and/or glucosyl-galactose moieties also is likely to affect fibrillar collagen organization [53]. Procollagen V is secreted and the C-propeptide is processed. However, unlike collagen I, the amino-terminal non-collagenous domain is only partially processed, with the cysteine-rich domain removed. An amino-terminal domain is retained containing a terminal tyrosine-rich, globular domain, a rigid rod-like domain and a flexible hinge region. This retained amino-terminal domain is responsible for most of the regulatory activity of the type V collagen molecule [54].

Collagens I and V collagen co-assemble so that the type V collagen triple helix is internalized within the fibril while the amino-terminal domain projects

through the gap region and is present on the fibril surface [49, 51]. The helical domain of collagen V is approximately 10% longer than the collagen I domain and does not perfectly fit a quarter-stagger arrangement with type I collagen [55, 56]. Properties intrinsic to these collagen-collagen interactions regulate the initiation and assembly of the initial fibril (intermediate). In situ experiments with corneal keratocytes demonstrated that reduction in the type V content increased fibril diameter [57]. In vitro self assembly assays using purified native type I and V collagen demonstrated that increasing the type V collagen content was inversely associated with fibril diameter. Removal of the N-terminal domain of type V collagen abolished most of this regulatory effect [49]. However, the co-polymerization of the type V triple helix with type I collagen retained a limited ability to influence fibril diameter. This effect may be related to the longer length of the type V triple helix relative to the type I helix and therefore less organized molecular packing. A strong relationship between regularity of molecular packing and larger aggregates has been shown in a number of systems [58, 59]. Therefore, the less regular packing possible with the heterotypic mixture would be related to smaller diameter fibrils. However, the N-terminal domain was required for most of the regulation of diameter. Molecular assembly of fibrils would require that this non-collagenous domain be present on the fibril surface. The presence of this domain on the fibril surface could limit fibril diameter via steric or electrostatic mechanisms. However, the quantitatively minor collagen may exert this regulatory influence by controlling nucleation of fibril assembly. The manipulation of collagen I/V ratios using stratified cultures of human cells haplo-insufficient in collagen V also demonstrated an inverse relationship between type V collagen content and fibril diameter [60]. Halving the collagen V decreased the initiation events; with a constant pool of type I collagen the result was assembly of fewer, larger diameter fibrils. This indicates that the interaction of collagen V with collagen I is required for the efficient initiation of fibril assembly. Therefore, the regulation of fibril diameter by collagen V is related to controlling the number of nucleation events. In the cornea, the high content of collagen V relative to other collagen I-containing extracellular matrices, i.e., 10–20% vs 1–2%, provide for the initiation of large numbers of small diameter fibrils. These immature fibrils are short intermediates forming the building blocks for the mature fibrils.

Collagen fibrils are initially formed as short fibril intermediates. Mature fibrils form via linear and lateral growth of the preformed intermediates [49, 61]. In cornea, transparency is dependent on the maintenance of small diameter fibrils and the mechanical properties require an increase in fibril length. The leucine-rich repeat proteoglycans interact with the fibril surface and have been implicated in the regulation of the later stages in collagen fibril growth. When corneal fibrils were isolated from the corneal stroma, stripped of fibril-associated molecules, increases in both length and diameter were observed [49]. This indicates that presence of the type V collagen N-terminal domain does not regulate later steps in fibril assembly. Decorin, lumican, keratocan and osteoglycine are all expressed in the corneal stroma and presumably all are bound

to fibril surfaces [14, 62–64]. This interaction of the proteoglycans with a fibril, forms a composite structure. This composite structure is important in the regulation of fibril growth, fibril packing and stromal hydration.

The proteoglycans found on the fibril surface function in maintaining the small homogenous fibril diameters necessary for transparency. Lumican-, keratocan- and osteoglycine-deficient mouse corneas all show increases in corneal fibril diameter [63–65]. However, only lumican-deficient corneas have major alterations in fibril and stromal structure associated with corneal opacity. Lumican-deficient mouse corneas have large irregular fibrils present in the stroma [62, 65]. These fibrils are indicative of a loss of diameter control. Since corneal fibrils grow in length, but not diameter, this indicates that the presence of lumican, as part of the collagen suprastructure in the cornea, is a necessary stabilizer/inhibitor of lateral fibril growth. The restriction of the major defects to the posterior stroma where lumican expression also is restricted in the wild type adult suggests that there are other comparable interactions involved in the anterior stroma. This indicates that different composites are found with different spatial distributions.

Transparency also depends on the rigid maintenance of stromal hydration. The charged proteoglycans bind water and keratan sulfate proteoglycans have essential roles in the regulation of corneal hydration. Another feature of the stroma necessary for function is the regular packing of stromal fibrils into regular orthogonal lattices. The surface associated proteoglycans have been implicated in this level of organization as well. In addition, type XII collagen, a FACIT collagen, is homogeneously expressed throughout the mature corneal stroma. The large non-collagenous domains of the fibril-associated molecules also have been implicated in the regulation of fibril packing.

5.3 Tendon Fibril Formation

In contrast to the cornea, the collagen fibrils in the tendon have significant increases in diameter during development and growth. The mature tendon contains uniaxial fibrils with a very heterogeneous population of different size fibrils. The mechanical properties of the tendon are dependent on the increases in fibril diameter seen with development. Tendon fibroblasts express collagen I as the quantitatively major fibril-forming collagen and minor amounts of collagens V and III. Both form heterotypic alloys with collagen I and have the retained/slowly processed N-terminal domains typical of the regulatory fibril-forming collagens. The nucleation of fibril assembly presumably proceeds as has been described in cornea, assembling relatively small diameter, short fibril intermediates.

During tendon development there are changing expression patterns for FACIT collagens and the leucine-rich proteoglycans [66–68]. Collagen XIV is expressed during early development followed by little if any expression. In the mouse, both biglycan and lumican are expressed at their highest levels during

development and then decrease rapidly to unmeasurable amounts with maturation. In contrast, expression of both decorin and fibromodulin increases during maturation and then remains stable. These changing expression patterns are reflected in the changing nature of the composite collagen fibrils during this period. Tendon fibril structures are abnormal in each strain of mice deficient in one of the four proteoglycans, consistent with abnormal regulation of fibril growth in the absence of any of the four proteoglycans [14, 69, 70]. These abnormalities are associated with alterations in the biomechanical strength of the tendons [69, 71]. The changing composite structure is therefore critical to the regulation of the linear and lateral growth steps necessary for normal tendon structure and therefore function.

The involvement of cellular structures in the deposition of tendon fibrils also has been the subject of several major papers [61]. Earlier investigations indicated that, the formation of primordial fibrils with a homogeneous diameter of ca. 30 nm occurred within deep recesses of tendon fibroblasts [72, 73]. This was confirmed in a recent report where intracellular processing of procollagen within elongated Golgi-to-plasma membrane compartments (GPCs) was demonstrated. Extrusion from the cells and deposition into pre-existing hexagonal arrays occurred through the formation of cellular protrusions, termed "fibripositors", that were formed by fusion of GPCs containing single or several small-diameter fibrils with the plasma membrane. Thus, a contiguity between novel organelles achieving fibril parallelism and extracellular fibril bundles was established. Fibripositors were prominent only during fetal development when fibril formation in tendons was at its peak, but were not observed during adolescence when tendon growth is the predominant event [74]. However, in analogy to sickling of erythrocytes containing deoxygenated hemoglobin S-fibers, generation of fibripositors may be a consequence of fibril self-assembly initiated intracellularly rather than a truly instructive process initiated by cytoskeletal structures and aligning parallel fibril bundles. In addition, the mechanism is not known whereby the primordial 30-nm fibrils being hexagonal cylindrical objects themselves are kept from fusing into large fibril bundles within the hexagonal packing into which they are arranged extracellularly. Since decorin-deficiency leads to abnormalities in fibrillar packing, it may well be that this or other small leucine-rich proteoglycans separate primordial fibrils from each other and tether them into the hexagonal arrays seen in channels between fetal tendon fibroblasts.

6 Concluding Remarks

There are at least 27 different collagens that can be grouped into subfamilies including: fibrillar collagens, FACIT collagens, and network-forming collagens that were the focus of this review. This diversity, alone, would generate numerous different collagen suprastructures. Increased complexity is obtained at

the level of alternative splicing and promoter use as well as post-translational modifications. In addition, the heteropolymeric nature of collagen suprastructures provides a mechanism for even greater diversity. The tissue-, domain- and developmental-specificity exhibited at all levels provides an extraordinary diversity in the building blocks of the extracellular matrix. A major challenge will be to elucidate the progressive changes in collagen suprastructures necessary during normal development as well as tissue repair and regeneration. A further understanding of the basis of connective tissue diseases also will require a definition of the interactions involved in assembly of tissue-specific suprastructures.

Acknowledgments Work from the authors' laboratories was supported by grants from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 492, project A2 and the National Institutes of Health, EY05129 and AR45745.

References

1. Prockop DJ, Kivirikko KI (1995) Annu Rev Biochem 64:403
2. Myllyharju J, Kivirikko KI (2004) Trends Genet 20:33
3. Myllyharju J, Kivirikko KI (2001) Ann Med 33:7
4. Gathercole LJ, Keller A (1991) Matrix 11:214
5. Keene DR, Lunstrum GP, Morris NP, Stoddard DW, Burgeson RE (1991) J Cell Biol 113:971
6. Niyibizi C, Eyre DR (1989) FEBS Lett 242:314
7. Brodsky B, Eikenberry EF, Belbruno KC, Sterling K (1982) Biopolymers 21:935
8. Hansen U, Bruckner P (2003) J Biol Chem 278:37352
9. Torre-Blanco A, Adachi E, Hojima Y, Wootton JA, Minor RR, Prockop DJ (1992) J Biol Chem 267:2650
10. Batge B, Winter C, Notbohm H, Acil Y, Brinckmann J, Muller PK (1997) J Biochem (Tokyo) 122:109
11. Notbohm H, Nokelainen M, Myllyharju J, Fietzek PP, Muller PK, Kivirikko KI (1999) J Biol Chem 274:8988
12. Keller H, Eikenberry EF, Winterhalter KH, Bruckner P (1985) Coll Relat Res 5:245
13. Holmes DF, Watson RB, Chapman JA, Kadler KE (1996) J Mol Biol 261:93
14. Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, Iozzo RV (1997) J Cell Biol 136:729
15. Eyre DR, Wu JJ, Fernandes RJ, Pietka TA, Weis MA (2002) Biochem Soc Trans 30:893
16. Vaughan L, Mendler M, Huber S, Bruckner P, Winterhalter KH, Irwin MI, Mayne R (1988) J Cell Biol 106:991
17. Koch M, Schulze J, Hansen U, Ashwodt T, Keene DR, Brunken WJ, Burgeson RE, Bruckner P, Bruckner-Tuderman L (2004) J Biol Chem 279:22514
18. Marneros AG, Keene DR, Hansen U, Fukai N, Moulton K, Goletz PL, Moiseyev G, Pawlyk BS, Halfter W, Dong S, Shibata M, Li T, Crouch RK, Bruckner P, Olsen BR (2004) EMBO J 23:89
19. Suzuki OT, Sertie AL, DerKaloustian V, Kok F, Carpenter M, Murray J, Czeizel AE, Kliemann SE, Rosemberg S, Monteiro M, Olsen BR, Passos-Bueno MR (2002) Am J Hum Genet 71:1320
20. Bruns RR, Press W, Engvall E, Timpl R, Gross J (1986) J Cell Biol 103:393

21. von der Mark H, Aumailley M, Wick G, Fleischmajer R, Timpl R (1984) Eur J Biochem 142:493
22. Kielty CM, Grant ME (2002) The collagen family: structure, assembly, and organization in the extracellular matrix. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders. Wiley-Liss, New York, p 159
23. Chu ML, Mann K, Deutzmann R, Pribula-Conway D, Hsu-Chen CC, Bernard MP, Timpl R (1987) Eur J Biochem 168:309
24. Ball S, Bella J, Kielty C, Shuttleworth A (2003) J Biol Chem 278:15326
25. Knupp C, Squire JM (2001) EMBO J 20:372
26. Ball SG, Baldock C, Kielty CM, Shuttleworth CA (2001) J Biol Chem 276:7422
27. Furthmayr H, Wiedemann H, Timpl R, Odermatt E, Engel J (1983) Biochem J 211:303
28. Baldock C, Sherratt MJ, Shuttleworth CA, Kielty CM (2003) J Mol Biol 330:297
29. Wiberg C, Hedbom E, Khairullina A, Lamande SR, Oldberg A, Timpl R, Morgelin M, Heinegard D (2001) J Biol Chem 276:18947
30. Wiberg C, Heinegard D, Wenglen C, Timpl R, Morgelin M (2002) J Biol Chem 277:49120
31. Wiberg C, Klatt AR, Wagener R, Paulsson M, Bateman JF, Heinegard D, Morgelin M (2003) J Biol Chem 278:37698
32. Yamaguchi N, Mayne R, Ninomiya Y (1991) J Biol Chem 266:4508
33. Shuttleworth CA (1997) Int J Biochem Cell Biol 29:1145
34. Sawada H, Konomi H, Hirosawa K (1990) J Cell Biol 110:219
35. Jakus MA (1956) J Biophys Biochem Cytol 2:243
36. Kwan AP, Cummings CE, Chapman JA, Grant ME (1991) J Cell Biol 114:597
37. Illidge C, Kielty C, Shuttleworth A (2001) Int J Biochem Cell Biol 33:521
38. Illidge C, Kielty C, Shuttleworth A (1998) J Biol Chem 273:22091
39. Stephan S, Sherratt MJ, Hodson N, Shuttleworth CA, Kielty CM (2004) J Biol Chem 279:21469
40. Kvansakul M, Bogin O, Hohenester E, Yayon A (2003) Matrix Biol 22:145
41. Eikenberry EF, Bruckner P (1999) Supramolecular structure of cartilage matrix. In: Shaw MK, Rom E, Bilezekian JP (eds) Dynamics of bone and cartilage metabolism. Academic Press, San Diego, CA, p 289
42. Poole CA (1992) Chondrons – the chondrocyte and its pericellular microenvironment. In: Kuettner KE, Schleyerbach R, Peyron JG, Hascall VC (eds) Articular cartilage and osteoarthritis. Raven Press, New York, p 210
43. Mendler M, Eich-Bender SG, Vaughan L, Winterhalter KH, Bruckner P (1989) J Cell Biol 108:191
44. Kassner A, Hansen U, Miosge N, Reinhardt DP, Aigner T, Bruckner-Tuderman L, Bruckner P, Grassel S (2003) Matrix Biol 22:131
45. Blaschke UK, Eikenberry EF, Hulmes DJ, Galla HJ, Bruckner P (2000) J Biol Chem 275:10370
46. Muller-Glauser W, Humbel B, Glatt M, Strauli P, Winterhalter KH, Bruckner P (1986) J Cell Biol 102:1931
47. Hagg R, Bruckner P, Hedbom E (1998) J Cell Biol 142:285
48. McLaughlin JS, Linsenmayer TF, Birk DE (1989) J Cell Sci 94:371
49. Birk DE (2001) Micron 32:223
50. Birk DE, Fitch JM, Babiarz JP, Linsenmayer TF (1988) J Cell Biol 106:999
51. Linsenmayer TF, Gibney E, Igoe F, Gordon MK, Fitch JM, Fessler LI, Birk DE (1993) J Cell Biol 121:1181
52. Birk DE, Lande MA (1981) Biochim Biophys Acta 670:362
53. Ibrahim J, Harding JJ (1989) Biochim Biophys Acta 992:9
54. Birk DE, Fitch JM, Babiarz JP, Doane KJ, Linsenmayer TF (1990) J Cell Sci 95:649

55. Fessler JH, Bachinger HP, Lunstrum G, Fessler LI (1982) Biosynthesis and processing of some procollagens. In: Kuehn K, Schoene H, Timpl R (eds) New trends in basement membrane research. Raven Press, New York, p 145
56. Silver FH, Birk DE (1984) Int J Biol Macromol 6:125
57. Marchant JK, Hahn RA, Linsenmayer TF, Birk DE (1996) J Cell Biol 135:1416
58. Makowski L, Magdoff-Fairchild B (1986) Science 234:1228
59. Weisel JW, Nagaswami C, Makowski L (1987) Proc Natl Acad Sci USA 84:8991
60. Wenstrup RJ, Florer JB, Cole WG, Willing MC, Birk DE (2004) J Cell Biochem 92:113
61. Birk DE, Linsenmayer TF (1994) In: Yurchenco PD, Birk DE, Mecham RP (eds) Extracellular matrix assembly and structure. Academic Press, NY, chap 4, p 91
62. Chakravarti S, Magnuson T, Lass JH, Jepsen KJ, LaMantia C, Carroll H (1998) J Cell Biol 141:1277
63. Liu CY, Birk DE, Hassell JR, Kane B, Kao WW (2003) J Biol Chem 278:21672
64. Tasheva ES, Koester A, Paulsen AQ, Garrett AS, Boyle DL, Davidson HJ, Song M, Fox N, Conrad GW (2002) Mol Vis 8:407
65. Chakravarti S, Petroll WM, Hassell JR, Jester JV, Lass JH, Paul J, Birk DE (2000) Invest Ophthalmol Vis Sci 41:3365
66. Ezura Y, Chakravarti S, Oldberg A, Chervonea I, Birk DE (2000) J Cell Biol 151:779–787
67. Young BB, Zhang G, Koch M, Birk DE (2002) J Cell Biochem 87:208
68. Zhang G, Young BB, Birk DE (2003) J Anat 202:411
69. Jepsen KJ, Wu F, Peragallo JH, Paul J, Roberts L, Ezura Y, Oldberg A, Birk DE, Chakravarti S (2002) J Biol Chem 277:35532
70. Young MF, Bi Y, Ameye L, Chen XD (2002) Glycoconj J 19:257
71. Robinson PS, Lin TW, Reynolds PR, Derwin KA, Iozzo RV, Soslowsky LJ (2004) J Biomech Eng 126:252
72. Birk DE, Trelstad RL (1986) J Cell Biol 103:231
73. Birk DE, Zycband EI, Winkelmann DA, Trelstad RL (1989) Proc Natl Acad Sci USA 86:4549
74. Canty EG, Lu Y, Meadows RS, Shaw MK, Holmes DF, Kadler KE (2004) J Cell Biol 165:553

Top Curr Chem (2005) 247: 207–229
DOI 10.1007/b103828
© Springer-Verlag Berlin Heidelberg 2005

Collagen Cross-Links

David R. Eyre (✉) · Jiann-Jiu Wu

Department of Orthopaedics and Sports Medicine, Orthopaedic Research Labs,
University of Washington, P.O. Box 356500, Seattle, WA 98195-6500, USA
deyre@u.washington.edu

1 Introduction 208

2 Fibril-Forming Collagens (Types I, II, III, V/XI) 208

3 Fibril-Associated Type IX Collagen of Cartilage 214

4 Basement Membrane Type IV Collagen 217

5 Lysyl Hydroxylases Regulate Tissue-Dependent Patterns of Cross-Linking . . . 218

6 Bone and Mineralized Tissue Collagens 220

7 Collagen Cross-Links as Biomarkers 221

8 Other Mechanisms of Collagen Cross-Linking 223
8.1 Cystine Disulfides 223
8.2 Gamma-Glutamyl Lysine Cross-Links 223
8.3 Tyrosine-Derived Cross-Links 224
8.4 Types VIII and X Collagens 224

9 Outlook 225

References 225

Abstract Collagen is the main source of extracellular support for multicellular animals. The mechanical strength of collagen fibrils depends on a highly regulated mechanism of intermolecular cross-linking. The basis of this cross-linking from the most primitive to the most advanced multicellular animals and across a diversity of vertebrate tissue types, is the formation of covalent bonds from aldehydes produced from lysyl and hydroxylysyl side-chains by lysyl oxidase. In the last decade it has become clear that such bonds form not only between collagen molecules of the same type in homopolymeric fibrils but also between different types of collagen molecule that have evolved to interact and form heteromeric structures. Furthermore, cross-linking amino acids and peptides containing them from collagen degradation, have received attention as bone resorption biomarkers in clinical studies and drug trials in the osteoporosis field. This review summarizes recent research directions with examples of advances in understanding complex interactions in cartilage collagen and the role of lysyl hydroxylase isoforms in regulating the pathway of cross-linking chemistry.

Keywords Extracellular matrix · Lysyl oxidase · Lysyl hydroxylases · Pyridinolines · Bone

1 Introduction

Collagen fibrils provide the mechanical support that enabled large multicellular animals to evolve on earth [1, 2]. The tensile strength of collagen depends on the formation of covalent intermolecular cross-links between the individual protein subunits [3]. All the fibril-forming collagen types in higher vertebrates (types I, II, III, V and XI) are cross-linked through a mechanism based on the reactions of aldehydes generated enzymically from lysine (or hydroxylysine) side-chains by lysyl oxidase [4, 5]. This pathway operates in the collagen fibrils of sponges (Porifera), the most primitive extant multicellular animals, through to mammals [6, 7]. Certain other collagen types (e.g., collagen type IX of cartilage) are also cross-linked by the lysyl oxidase mechanism. In the last decade, perhaps the most important conceptual advance in this field is that different collagen types have evolved to cross-link heterotypically in the assembly of multi-component fibrils. Templates of one class of collagen (type V/XI microfibrils) provide the scaffold on which the more recently evolved fibrillar collagen molecules (types I and II) co-assemble to form large fibrils. In cartilage, a third type of collagen, type IX (of the FACIT sub-family), becomes cross-linked to the surface of this copolymer and all three molecular types are cross-linked internally and to each other through the lysyl oxidase mechanism. In considering the evolution of the different molecular classes of collagen, it is clear that gain of function mutations in the protein subunits have created new cross-linking opportunities for heterotypic intermolecular bonding. This presumably extends the range of collagen functional properties cells can express.

In addition to physiologically regulated cross-linking on synthesis by lysyl oxidase, collagens are also susceptible over time to further cross-linking through the undesirable reactions of reducing sugars [8–13], particularly glucose (non-enzymic glycosylation or glycation) and lipid oxidation products [14, 15]. There is an extensive and growing literature on the chemistry of such age-related changes, which in general have pathological effects, but this chemistry will not be reviewed here.

2 Fibril-Forming Collagens (Types I, II, III, V/XI)

Most of the research defining the cross-linking pathways from precursor lysine and hydroxylysine aldehydes in bulk collagen fibrils of major connective tissues was done over 20 years ago (reviewed in [4, 16–19]). The basic pathways are shown in Figs. 1 and 2. In the last decade the prominence of pyrrole cross-links as maturation products in bone collagen and their molecular location have been established [20].

The two basic pathways, lysine aldehyde vs hydroxylysine aldehyde-initiated, appear in general to occur in loose vs stiff connective tissues respectively. In

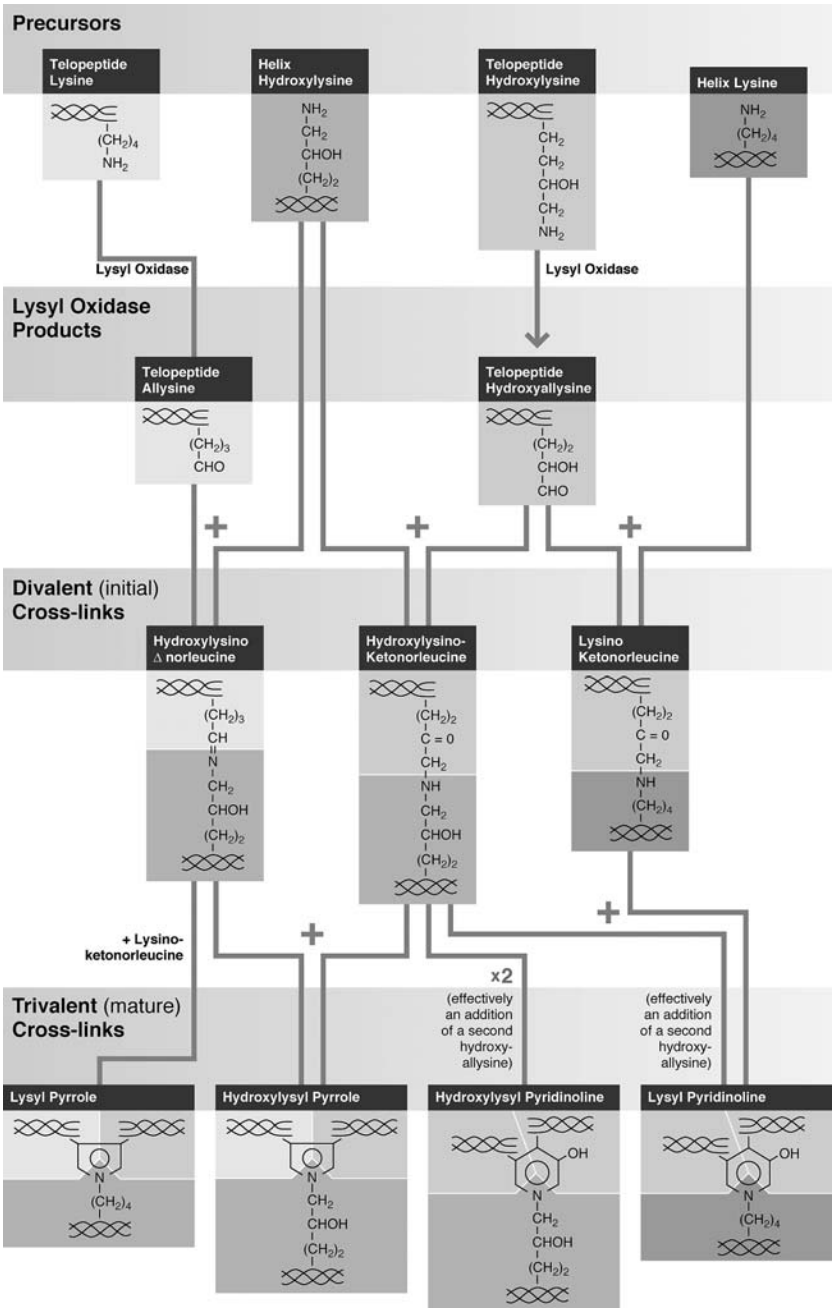

Fig. 1 The hydroxyallysine cross-linking pathway. Hydroxylysine residues are the source of aldehydes formed by lysyl oxidase for intermolecular cross-linking reactions. Mature cross-links are trivalent pyridinolines. Bone collagen is unusual in that pyrrole cross-links are also prominent because both hydroxyallysine and allysine participate as precursors

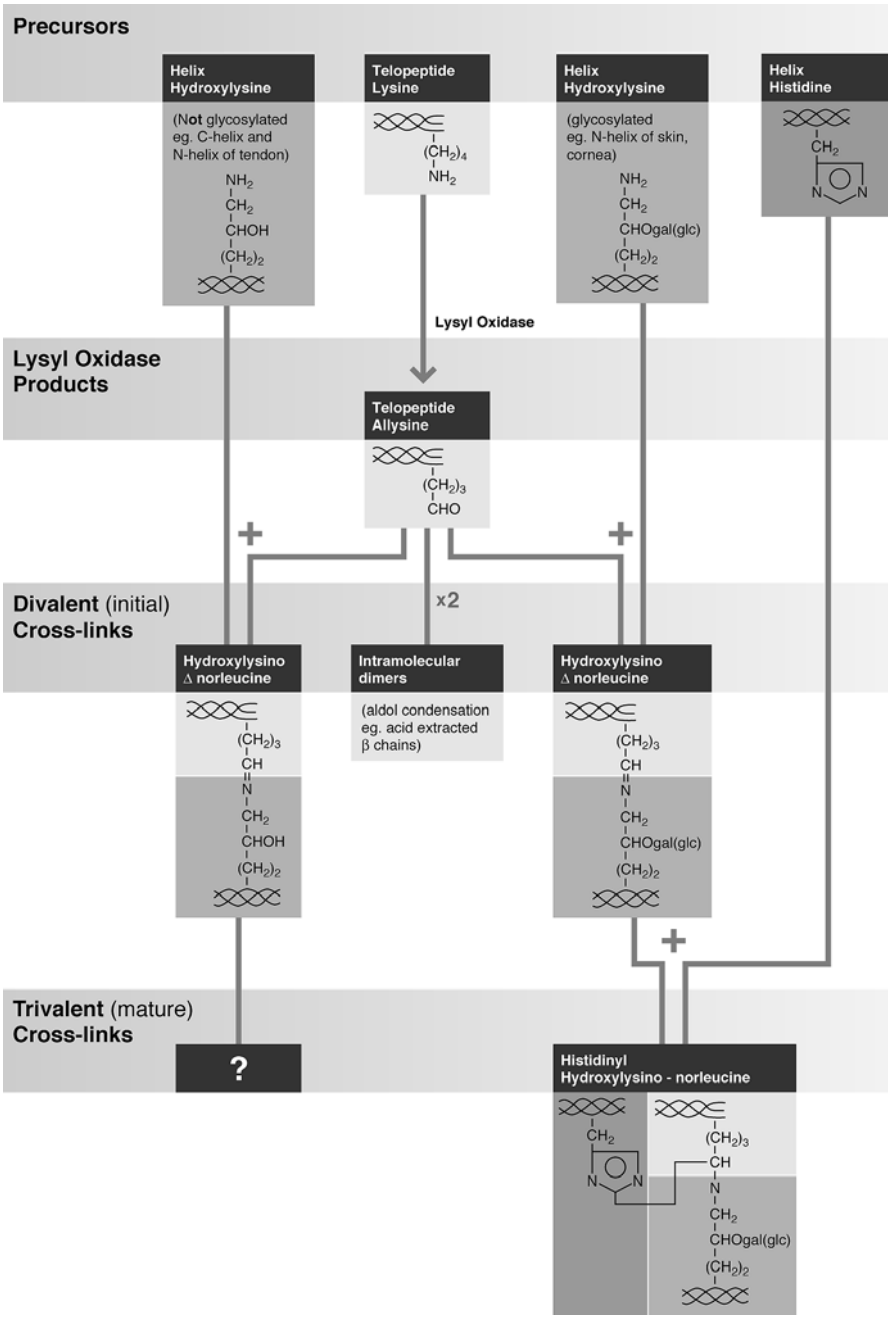

Fig. 2 The allysine cross-linking pathway. Lysine residues are the source of aldehydes formed by lysyl oxidase for intermolecular cross-linking reactions. Histidine can participate in mature cross-link formation, notably in skin collagen

detail, the tissue specificities and the pathways are more diversified, with elements of both pathways and both precursor aldehydes combined in some specialized tissues, for example in bone type I collagen [20]. Historically, the lysine aldehyde pathway (Fig. 2) was easier to study since the initially-formed aldimines are cleaved at low pH allowing the collagen monomers to be solubilized from rat skin and tail tendon in 0.5 M acetic acid [19]. On the hydroxylysine aldehyde pathway (Fig. 1), neither the initial cross-links nor their maturation products were labile so these collagens were insoluble making molecular analysis more difficult. The natural fluorescence of the pyridinoline cross-links (Fig. 1) allowed peptides to be isolated and their molecular sites of origin to be worked out [21–23]. In the last decade the greatest attention in the literature to pyridinoline residues, and to collagen cross-links in general, is from publications reporting assays of these residues and peptides containing them in blood and urine as biomarkers of bone resorption and connective tissue degradation. Because pyridinolines cannot be metabolized, their levels in blood and urine provide a measure of the amount of collagen and hence tissue from which they were proteolytically derived.

Variations in cross-linking chemistry appear to be more tissue-specific than collagen type-specific. This is understandable if a single cell type synthesizing multiple collagens passes them through the same processing enzymes and endoplasmic reticular pathway. The basic pathway of cross-linking is regulated primarily by the hydroxylation pattern of telopeptide and triple-helix domain lysine residues. The flanking sequences around the cross-linking lysine residues, however, can affect the ensuing chemistry. An example is the participation of a histidine residue in the formation of the mature trivalent cross-links found in skin type I collagen (Fig. 2). A histidine in the α2(I) chain (of a third collagen molecule) reacts with a vicinal aldimine cross-link formed between a lysine aldehyde and a hydroxylysine residue in two 4D-staggered collagen I molecules [24, 25]. The pitch of collagen molecules packed in skin collagen fibrils is thought to facilitate this addition reaction to histidine [26], but whether it is fortuitous or provides a functional advantage to dermal collagen is unknown. Older studies had observed that the helical domain hydroxylysine at this location in the α1(I) chain was glycosylated in skin, but not in tendon [27]. Notably, the histidine-containing mature cross-link HHL (histidinyl hydroxylysino-norleucine), is not found in tendon collagen, which raises the possibility that glycosylation might drive HHL formation. It has been observed that pyridinoline residues (hydroxylysyl pyridinoline) can still be formed when glycosylated hydroxylysine occurs in the helix, for example at the C-telopeptide to N-helix site in bone collagen [20]. But at the N-telopeptide to C-helix cross-linking site, the helical hydroxylysine is not glycosylated. Again the biological significance of these differences is unknown.

Edman N-terminal sequence analysis of cross-linked peptides isolated from protease digests of tissue collagens has revealed intermolecular cross-linking between types I and III collagens (e.g., from aorta [28, 29]) and types I and II collagens (e.g., from intervertebral disc [30]). Unequivocally proving covalent

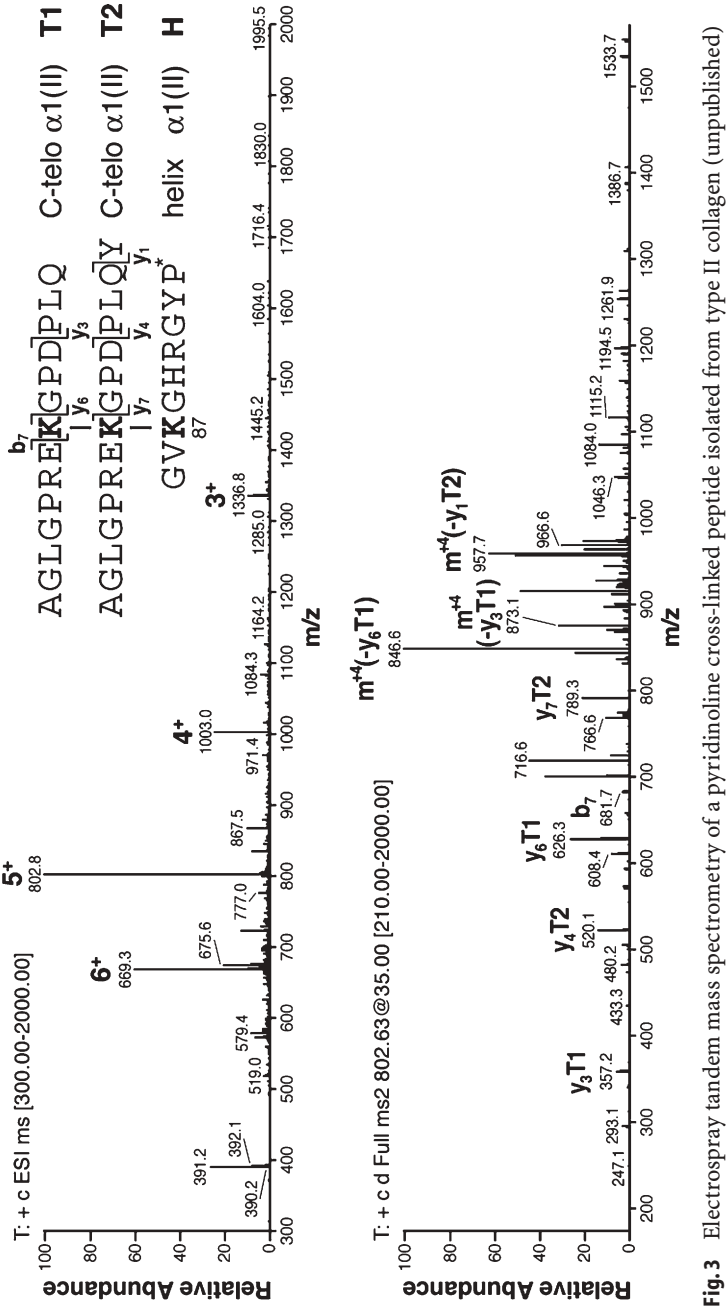

Fig. 3 Electrospray tandem mass spectrometry of a pyridinoline cross-linked peptide isolated from type II collagen (unpublished)

structures rather than peptide mixtures is difficult using chromatography and N-terminal sequence analysis alone. Mass spectrometry promises to be a valuable tool in such studies. Preliminary data from purified cross-linked trimeric peptides indicate characteristic MS/MS fragmentation patterns. Figure 3 shows an example of MS/MS data from tandem mass spectrometry of a homotypic fragment from human cartilage type II collagen.

Types V and XI collagens are quantitatively minor molecular species that are found copolymerized with collagens I and II respectively, and are believed to form a filamentous scaffold or template on which the bulk fibrillar collagens co-polymerize. The gene products represented in these molecules can in fact assemble in a variety of heterotrimeric chain combinations, not just type V ($[\alpha1(V)]_2\alpha2(V)$) and type XI ($[\alpha1(XI)]$ $[\alpha2(XI)]$ $[\alpha3(XI)]$) molecules, but also novel tissue-specific chain associations [31–33]. Collagen type V/XI is best considered as a distinct sub-family of the fibril-forming collagen molecules. The same basic pattern of cross-linking, employing lysyl oxidase, and the telopeptide-to-helix location of intermolecular bonds occurs in these molecules. Homologous cross-linking sites (lysines) to those in types I, II and III collagens are evident in the protein sequences (Fig. 4), consistent with their evolutionary origin from a single founder gene [2].

Analyses of cross-linked peptides isolated from type XI collagen of cartilage showed only divalent cross-links (hydroxylysino-5-*keto*-norleucine), not the mature form of pyridinoline (hydroxylysyl pyridinoline) that predominates in type II collagen, the bulk fibril-forming subunit of the copolymer [34]. In addition, analysis of cross-linked peptides showed that most of the intermolecular bonds had formed between collagen XI molecules (N-telopeptide to C-helix). Any inter-type cross-linking was between type II C-telopeptides and the type XI N-helical site. This is best interpreted if collagen XI had self-polymerized to form its own filamentous network, at least initially, consistent with its suspected role as a template for the growth of thick banded fibrils. Similar properties have been

Type	Chain	N-telo	N-helix	C-helix	C-telo
I	α1(I)	YDEKSTGGI	GMKGHR	GIKGHR	PPQEKAHDG
	α2(I)	YDGKGVGLG	GFKGIR	GLKGHN	no lysine
II	α1(II)	FDEKAGGAQ	GVKGHR	GLKGHR	GPREKGPDP
III	α1(III)	YDVKSGVAV	GMKGHR	GIKGHR	IGGEKAGGF
V/XI	α1(V)	AGSKGPMVS	GEKGHR	GEKGHP	no lysine
	α2(V)	LDEKSGLGS	GLKGHR	GQKGHR	no lysine
	α3(V)	GSFKGPPVS	GEKGQR	GEKGHI	no lysine
	α1(XI)	DGSKGPTIS	GDKGHR	GEKGHP	PILSSKKTRR[1]
	α2(XI)	GGDKGPVVA	GEKGHR	GEKGHP	PIQMPKKTRR[1]

1. Lysines appear not to be involved in cross-linking

Fig. 4 Amino acid sequences at the four primary cross-linking sites in fibrillar collagens. Related sequence motifs can be found in all the fibrillar collagen gene products

found for type V collagen isolated from bone [35]. The primary cross-links were divalent (hydroxylysino-5-*keto*-norleucine) and they linked type V molecules between N-telopeptides and the C-helix site in a head-to-tail manner. Cross-links from the C-telopeptide of type I collagen to the N-helix site of type V were also found. The data on mature bone fit a special role for a hybrid type V/XI collagen molecule, [α1(V)α1(XI)α2(V)] [32], as a component of the filamentous template of the type I collagen-based fibrillar network of bone matrix.

3 Fibril-Associated Type IX Collagen of Cartilage

In addition to the fibril-forming collagen molecules (and basement membrane type IV collagen), 30 or more other molecular forms of collagen function in the extracellular matrix or at cell surfaces. Of these, only type IX collagen has been established to use the lysyl oxidase-mediated pathway of cross-linking. Within the diverse FACIT sub-family of collagen molecules (fibril-associated collagen with interrupted triple-helix [36]) therefore, collagen IX is the only member known to cross-link by the lysyl oxidase mechanism. Collagen IX cross-linking is extensive, highly evolved and presumably central to the protein's function as an adapter molecule on the surface of nascent type II collagen fibrils. All three collagen IX chains, which in birds and mammals are each the product of a distinct gene, appear to have diverged from a common ancestor after earlier whole or partial genome duplications. Each chain has two or three sites through which lysine-mediated cross-links can form and each displays unique specificity in its evolved cross-linking interactions. Through a series of studies over the last decade, we have defined the cross-linking sites and nature of the cross-linking of collagen type IX [37–41] (Fig. 5). The experimental approach has focused on isolating and structurally identifying peptides containing cross-linking residues (using fluorescence to track pyridinolines, and NaB^3H_4 to stabilize and tritium-label divalent cross-links). Conventional Edman-sequencing and, more recently, liquid chromatography/electrospray mass spectrometry (LCMS [41]) have been the methods of choice. LCMS is clearly a powerful tool for the future for defining complex cross-linking mechanisms in collagens and other extracellular matrix proteins.

Collagen IX has evolved to cross-link to collagen type II. The positioning of cross-linking sites along the molecule predicts their spatial inter-relationship. Figure 6 shows a molecular interaction model that can accommodate all six known cross-linking sites, and their interaction partner residues in type II collagen and other type IX collagen molecules. This packing arrangement requires an antiparallel relationship between the central type IX COL2 triple-helical domain and type II collagen molecules on a fibril surface and the type IX COL1 domain to be folded back on the COL2 domain as shown. All the known bonds [22, 37, 38, 41] can then be accommodated with the correct axial spatial alignments for cross-link formation.

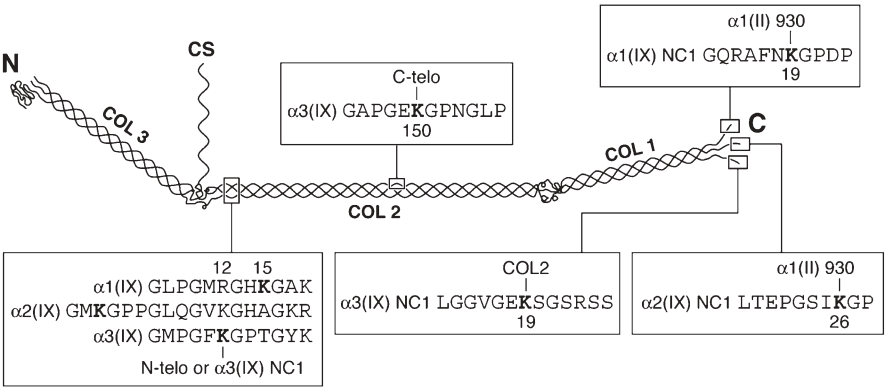

Fig. 5 Cross-linking sites in type IX collagen. Seven sites (hydroxylysine or lysine residues) in total have been identified, two each in α1(IX) and α2(IX) and three in α3(IX)

Clearly it will be important to define the molecular interactions that regulate this complex cross-linking cascade, which also includes the participation of collagen type XI, acting as a template filament within the composite fibril. Are monomers of collagen IX transported extracellularly before they interact with nascent (thin) collagen II/XI co-assemblies, or is a discrete element of the collagen II/IX/XI heteromer pre-fabricated inside the cell in a secretory compartment, which can polymerize extracellularly with the addition of type II collagen monomers? It is known that collagens IX and XI are most concentrated (with their highest ratios to collagen type II) in developing cartilage and in the immediate pericellular zone of chondrocytes [42–44]. As cartilage matures, the ratios of collagen II to collagens XI and IX rise to about 96:3:1 from 80:10:10 in fetal cartilage [39].

What is the function of collagen IX? Why are covalent bonds needed to anchor it to the surface of collagen II fibrils and to link adjacent collagen IX molecules? A mechanical role in strengthening the fibrillar matrix is reasonable to suspect. The interaction scheme shown in Fig. 6 suggests how interfibrillar cross-links might also form and so add to network stability, but the existence of such bonding will be hard to rule in or out. It does appear that the presence of collagen IX is not crucial for skeletal growth, since mice engineered with homozygous null genes for COL9A1, in which type IX collagen is functionally knocked out [45], appear normal at birth [45, 46]. They do, however, develop osteoarthritis with progressive destruction of joint cartilages [46], as do mice and humans expressing mutated forms of collagen IX genes [47, 48]. Collagen IX therefore may be required for mechanically durable mature cartilages. Perhaps the three-dimensional meshwork of developing fibrils is locked as a template outside the cell through the covalent interactions of collagen IX. Without a covalently-stabilized template, the mature collagen network may be more susceptible to proteolysis, detrimental effects of mechanical loads and less able to endure repetitive injury and repair cycles.

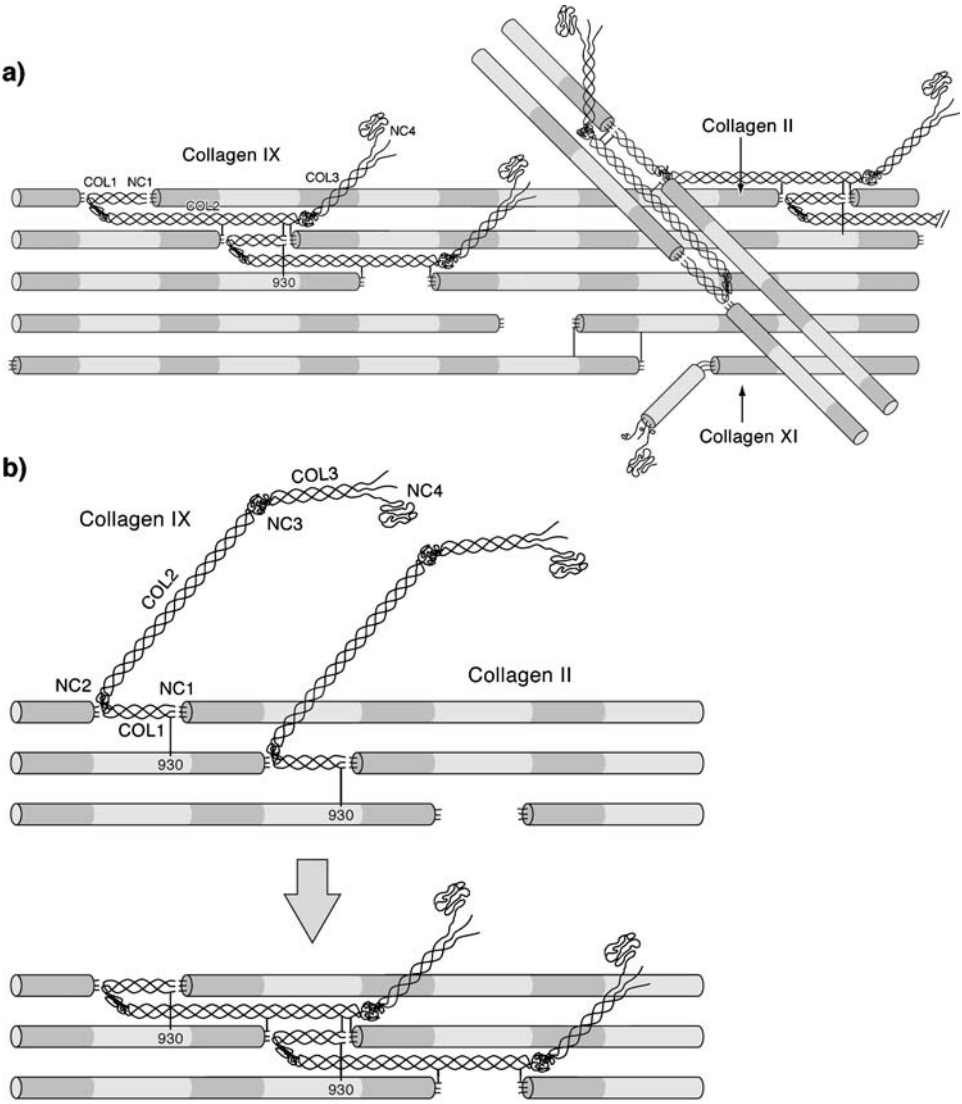

Fig. 6 A proposed model of interaction of type IX collagen with type II collagen that can accommodate all the cross-linking sites shown in Fig. 5. Folding back of the type IX COL1 domain on COL2 and an antiparallel alignment of the molecular on the surface of the type II collagen polymer can accommodate all the cross-linking interactions in the type IX collagen molecule. Potential inter-fibrillar bonds are also illustrated. The type II collagen fibril surface is shown acting as a template to bind and position adjacent type IX collagen molecules to cross-link to each other

The complex structure of the cartilage heterofibril network presents a challenge for understanding how the various enzyme-controlled steps in fibril growth are orchestrated. Lysyl oxidase must continue to create aldehydes on multiple sites on the interacting structural molecules without interfering spatially with the addition of new subunits or the propeptidases that remove the globular propeptides from II and XI procollagen molecules. The procession of protein interactions driving these and other activities on the fibril surface will be important to understand.

Inspection of the gene sequences for the other FACIT family members reveals no obvious homologies to suggest candidate cross-linking lysines in domains comparable to those in type IX collagen. One possible exception is COL21A1, which has a candidate lysine in its C-terminal NC1 domain, which also is short as in the three type IX collagen chains [49]. Little is yet known about the tissue distribution and properties of collagen XXI, for example whether it binds to collagen fibrils and so could act in a similar manner to collagen IX. It does not appear to be present in cartilage.

4
Basement Membrane Type IV Collagen

Basement membrane collagens are ancient (>500 million years [50, 51]), having evolved in primitive metazoa as early or earlier than the fibril-forming collagens. Hydra, a simple organism formed from two cell layers that secrete and sandwich the mesoglea, an extracellular layer, has been shown to express genes for a basement membrane collagen that is homologous in sequence to vertebrate type IV collagen and for a fibril-forming collagen [52, 53].

The open molecular networks that collagen type IV molecules form [54, 55] provide the framework for an assortment of proteins and proteoglycans that characterize the basement laminae that underlie most endothelial and epithelial cell layers. Early work indicated that vertebrate type IV collagen molecules are cross-linked covalently by disulfide bonds and bonds derived through the aldehyde-initiated lysyl oxidase mechanism [56]. Lysyl oxidase-mediated cross-links and disulfides were found in the 7S domain, a cross-linked tetramer of N-terminal globular domains, and also in preparations of the dimeric C-terminal NC1 domains. Recent work has concluded that the apparent non-reducible cross-link (suggesting a lysyl oxidase product) in the NC-1 dimer, was in fact a stable disulfide bond that required a high concentration of mercaptoethanol to break [57]. On the other hand, a high-resolution X-ray crystallographic study has indicated a strong spatial association between a methionine and a lysine side-chain in this domain [58], suggesting a novel covalent intermolecular bond although the chemical nature of the link was not determined. Whether lysyl oxidase-mediated cross-links ever occur in type IV collagen still seems to be an open question. Both divalent cross-links (hydroxylysino-5-*keto*-norleucine; [56]) and evidence for pyrroles [59] have earlier been noted,

but the low yield of divalent cross-links implied that the main cross-links remained unidentified.

An alternative explanation is that the low levels of known collagen cross-links had come from other collagen contaminants rather than from type IV collagen itself, and that in fact type IV collagen networks are not cross-linked by the lysyl oxidase mechanism. To resolve this question, more definitive data on cross-linked purified peptides are needed. On current evidence it seems that cystine disulfides and strong hydrophobic interactions may explain the cross-linking properties of type IV collagen [55]. The collagen type IV gene product of *Hydra vulgaris* lacks a 7S domain in its sequence by comparison with the six vertebrate collagen IV genes, and the stable non-reducible cross-links recently noted in its NC1 domain [52] could be an unusually stable disulfide, as recently concluded for this domain from vertebrate type IV collagen [57]. The highly conserved cysteine distribution patterns in the type IV collagens of Hydra and vertebrates support this possibility [52].

5 Lysyl Hydroxylases Regulate Tissue-Dependent Patterns of Cross-Linking

From the structures of the lysyl oxidase-mediated cross-links and tissue-specific differences, it has long been suspected that one or more telopeptide lysyl hydroxylases must exist to regulate the different cross-linking pathways. Only recently has direct evidence for such a gene product been found. Bank and colleagues discovered that the defect in Brück Syndrome, a heritable disorder of bone resembling osteogenesis imperfecta, eliminated pyridinolines and other hydroxylysine aldehyde-based cross-links from bone collagen [60], implying a defect in the putative telopeptide hydroxylase. In the first family studied, it turned out that the chromosomal locus linked to disease expression was not the hydroxylase, and the effect was probably through an associated gene product. This was revealed later in a study of two other families expressing the Brück Syndrome phenotype and the same abnormal pattern of bone collagen cross-linking, in which disease-causing mutations were identified in PLOD2 (procollagen-lysine 2-oxoglutarate 5-dioxygenase, also known as lysyl hydroxylase 2 or LH2) [61], one of the three human genes that encode lysyl hydroxylase isoforms [62–65]. PLOD2 can be expressed in two alternative splicing variants, LH2a and LH2b, where LH2b contains the product of an extra exon [66]. Studies on skin fibroblasts from patients with systemic sclerosis, which features skin progressively fibrotic in which the collagen had higher than normal levels of hydroxylysine-aldehyde cross-links, showed overexpression of LH2b, which is concluded to be the telopeptide hydroxylase [61].

The molecular basis of another genetic disease, Ehlers-Danlos Syndrome type VI (EDS-VIA) caused by mutations in PLOD1 (LH1) [67–69], have helped in understanding how the chemical quality of collagen cross-linking is con-

trolled in vivo. In EDS-VIA, active PLOD1 expression is eliminated and in bone collagen, HP (hydroxylysyl pyridinoline), is replaced by LP (lysyl pyridinoline), which means that the specific triple-helical cross-linking lysines that donate the ring-nitrogen side-chain of pyridinolines are underhydroxylated [70, 71]. In addition, skin type I collagen of EDS-VI patients essentially lacks any hydroxylysine residues, whereas bone type I collagen has about 50% of normal hydroxylysine. EDS-VIA cartilage type II collagen has 90% of the hydroxylysine content of normal cartilage but its HP/LP ratio is abnormally low [71]. Together these findings reveal tissue-specific and molecular site-specific differences in the relative contributions of LH1, LH2 and LH3 to triple-helical domain lysine hydroxylation. Also, the product of PLOD1 (LH1) is a helical lysyl hydroxylase that favors lysine residues as substrates at the two triple-helical sites of cross-linking in fibrillar collagens. In considering how these hydroxylases may regulate cross-linking, another property that may be important is the apparently additional activities of LH3 as both a galactosyl transferase and glucosyl transferase for collagen [72–74]. Since glycosylated hydroxylysines at cross-linking sites can participate in cross-linking, this may turn out to be another regulatory step. Notably in EDS VIA in skin collagen lacking any hydroxylysine, the cross-link later identified as HHL (Fig. 2) was missing (Eyre, unpublished). This suggests that lysine cannot substitute for glycosylated hydroxylysine (which is the helix donor residue) and form the lysine homologue (see Fig. 2).

In bone collagen the lysyl hydroxylase-mediated control mechanisms for cross-linking must be especially fine-tuned. Each triple-helical domain lysine and telopeptide lysine that goes on to form cross-links is partially hydroxylated in a site-specific manner. This produces the characteristic ratio of HP/LP and of pyridinolines to pyrroles that typify bone collagen. Presumably, expression levels of lysyl hydroxylase 1, 2 and 3 by osteoblasts and the local sequence context of the individual chains in which the cross-linking lysines occur, dictate the pattern. Analysis of the eventual cross-link composition of peptides from each locus can be used to estimate the original degrees of lysine hydroxylation [20]. Figure 7

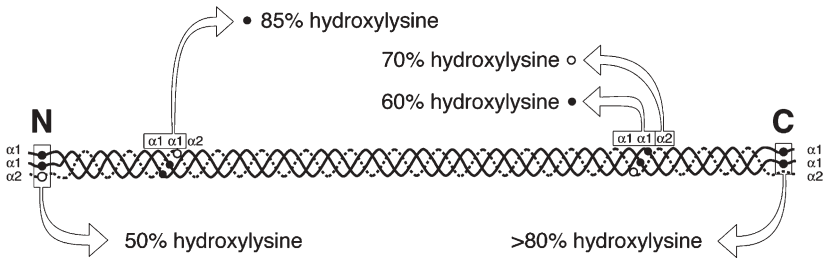

Fig. 7 A pattern of partial hydroxylation of cross-linking lysine residues in bone collagen regulates this tissue's distinctive cross-linking chemistry. The cross-link properties and yields of cross-linked peptides from the four primary loci are the source for the degrees of hydroxylation indicated [20]

summarizes the information from such analyses of normal human bone collagen. How this affects or relates to the lateral organization of molecules in bone collagen fibrils and the process of mineral crystallite deposition is still not clear, though suspected to be fundamental to bone properties [75]. In a study of LH1, LH2 and LH3 in the rat, most tissues expressed mRNAs for all three enzymes, implying a lack of tissue specificity in lysyl hydroxylase function [76]. However, collagen site-specific hydroxylation is probably regulated at more than one level of gene and protein expression.

6 Bone and Mineralized Tissue Collagens

A surge of interest in the last decade in collagen cross-links as clinical biomarkers of bone turnover has drawn attention to the cross-linking properties of bone collagen. It has long been noted that the cross-linking of collagen in bone, and other tissues that mineralize (dentin, calcified tendons) is distinctive, and probably functionally related to the intimate relationship between mineral crystallites and the packing arrangement of collagen molecules in fibrils [77, 78]. The mechanism, using both lysine aldehydes and hydroxylysine aldehydes, is distinctive and produces roughly equal amounts of pyridinolines and pyrroles as the mature cross-linking residues (see Fig. 1). Each type of cross-link is distributed site-specifically. Pyrroles, for example, are concentrated at the N-telopeptide-to-α2(I) chain C-helix [20]. Lysyl pyridinoline is also more abundant in general at the N-telopeptide-to-C-helix locus.

The structural quality of the trabecular architecture of human cancellous bone was found to be linked to the ratio of pyrrole to pyridinoline cross-links [79]. Also, a change in the cross-linking pattern along turkey tendons just before they mineralize suggests a causative relationship between cross-linking quality and mineralization [80]. In a fracture repair model, the ratio of HP/LP and hydroxylysine content of the callus collagen was strongly related to the degree of fibril mineralization [81]. Methods for detecting cross-linking residues spectroscopically in sections of bone tissue using FTIR show promise in exploring this further [82–84]. Pyrrole cross-links, which are not stable to acid hydrolysis, are difficult to quantify and characterize. A method that uses biotinylated Ehrlich's reagent to derivatize the residues prior to isolation has been reported [85]. Lysyl pyridinoline, a recognized biomarker of bone collagen, has now been synthesized as a standard and reagent source for developing biomarker immunoassays [86].

A further form of pyrrole cross-link, a trivalent pyrroleninone, has been tentatively identified in an acid hydrolysate of dentin collagen (Fig. 8) [87], in addition to the pyridinolines, pyrroles and ketoamines already described.

In another study using an antibody that recognizes the C-telopeptide of the α1(I) chain, both cross-linked and uncross-linked forms of this domain were isolated from trypsin-digested human bone collagen [88]. Of the trivalent

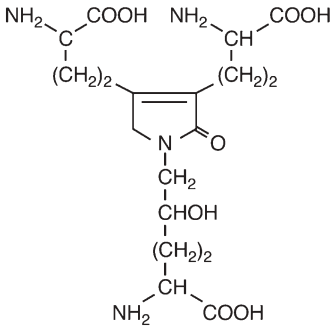

Fig. 8 A novel pyrroleninone cross-link isolated from bovine dentin in addition to pyridinolines, pyrroles and divalent keto-amines

cross-linked structures, a significant fraction could not be explained by pyrroles or pyridinolines, implying an unidentified cross-linking residue.

Higher levels of pyridinoline cross-links and hydroxylysine were found in bone from patients with osteogenesis imperfecta (clinical types I, II and III) compared with normal bone [89]. Although this suggests no gross disturbance in molecular packing of collagen molecules in fibrils, the potential that collagen/mineral inter-relationships are subtly disturbed deserves more attention as a source of the bone fragility.

7 Collagen Cross-Links as Biomarkers

Results from searching the literature on collagen cross-linking are heavily weighted over the last decade with reports from clinical studies that measured pyridinolines or cross-linked telopeptides in body fluids as molecular markers of bone turnover (reviewed in [90–94]). These cross-linking amino acids, and peptides containing them, are found in blood as products of collagen proteolysis which survive into urine. Pyridinolines are quantitatively excreted in the form of the free amino acids and small peptides, there being no degradation pathway in the liver. Initial reports showed that the total pool of pyridinolines (HP plus LP) in urine, quantified after acid hydrolysis by HPLC, can provide a more specific index of systemic bone resorption than hydroxyproline, a long-used marker [95–99]. Even more specific immunoassays were then introduced that targeted short telopeptide fragments attached to the cross-linking residues using specific antibodies [100–102]. It was discovered that peptide degradation products of the two cross-linked domains of bone collagen (N-telopeptide-to-helix or NTx, and C-telopeptide-to-helix or CTx) surviving into blood and urine, fell into discrete chromatographic pools of low molecular weight (<2 kDa) [100–103]. Antibodies could be tailored that recognized the core peptide components as neoepitopes. Immunoassays were developed that are specific to type I

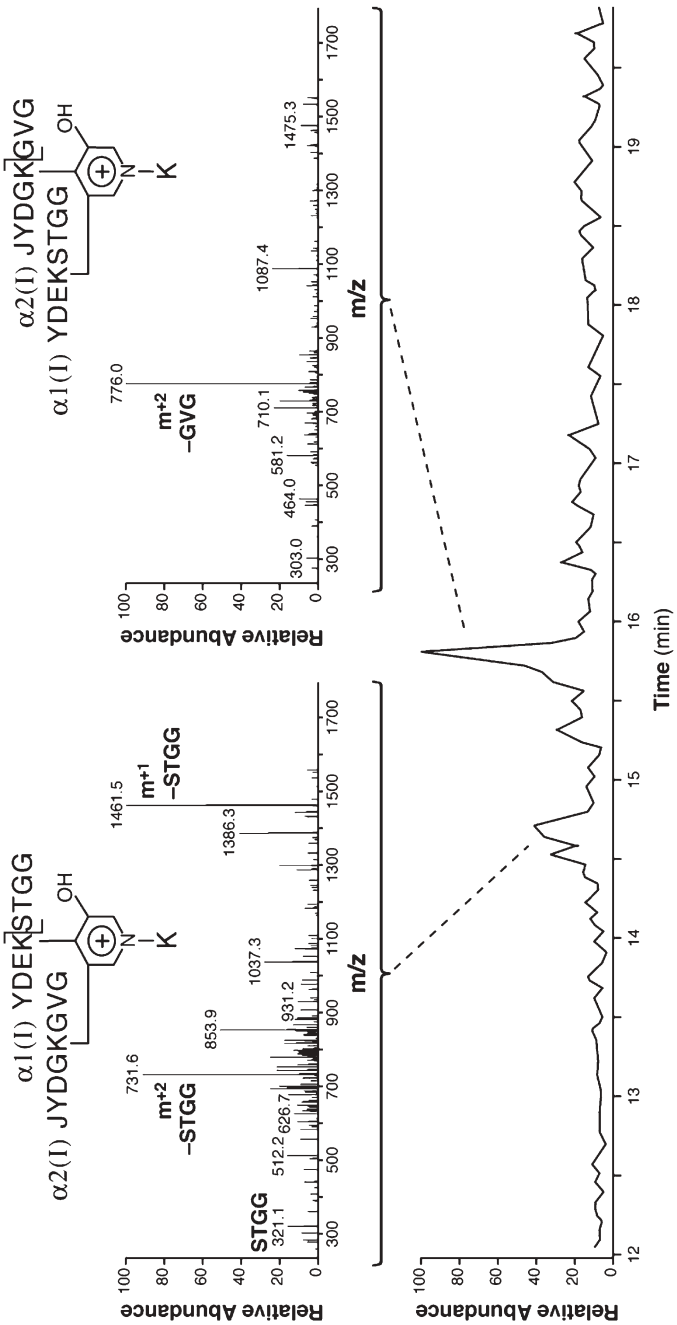

Fig. 9 Stereoisomers of a cross-linked type I collagen peptide (NTx) from human urine resolved by LCMS (liquid chromatography/mass spectrometry). The peptides differ in having the pyridinium ring positions of their telopeptide donor arms reversed. The ring nitrogen donor peptide sequence has lost all attached amino acids (presumably through proteolysis in the kidney before excretion)

collagen (peptide sequence specificity) and to cleavage products peculiar to the cathepsin K-initiated pathway of proteolysis which osteoclasts use to degrade bone collagen [104, 105].

Figure 9 shows an NTx peptide structure recovered from urine. Using in-line microbore reverse-phase HPLC and electrospray mass spectrometry, two stereoisomers of the peptide can be resolved, which differ in MS/MS fragmentation pattern. This can be explained by an interchanged placing of their N-telopeptide side-arms (α1(I) and α2(I)) on the 3-hydroxy-pyridinium ring of the trivalent cross-linking residue as shown. Mass spectrometry reveals a favored primary fragmentation product on MS/MS that we presume depends on the pyridinium ring positions of the peptide arms. This information can help in defining the origin and preferred path of interactions of the three precursor collagen sequences giving rise to each trivalent structure. Liquid chromatography/mass spectrometry also resolved the HP and LP forms of the cross-links (Fig. 9 shows the spectra for the LP stereoisomers).

Although bone collagen is the principal source of pyridinoline cross-links in urine, other tissue collagens also contribute. For example, specific fragments from cartilage type II collagen have been identified [106] and targeted for immunoassay as a biomarker of cartilage breakdown [107]. The main source in urine of these discrete peptides from type II collagen we believe is from osteoclastic breakdown of mineralized cartilage by osteoclasts. Levels are extremely high in growing children with open growth plates, supporting this conclusion [108]. Patients with osteoarthritis show on average higher levels than control subjects [109] and, in a study of high-performance athletes, runners showed higher levels than swimmers or rowers [110]. Accelerated joint remodeling is the likely explanation for the raised levels in adults.

8
Other Mechanisms of Collagen Cross-Linking

8.1
Cystine Disulfides

For most non-fibrillar collagens of the extracellular matrix (e.g., type IV collagen, see earlier) cystine cross-links may be the only source of covalent intra- and inter-molecular bonds. Type VI collagen forms a characteristic banded filamentous network in which the chains are cross-linked as molecules, dimers and tetramers by disulfides. No lysine-mediated or other cross-links are present [111].

8.2
Gamma-Glutamyl Lysine Cross-Links

In addition to lysyl oxidase-mediated cross-links, there is indirect evidence that transglutaminase-mediated cross-links might also be involved in the process

of polymerization of collagen networks [112–114]. Candidate glutamine substrate sites have been identified for transglutaminase, but no partnering lysine residues for natural cross-link formation have been defined.

Tissue transglutaminases, related to Factor XIII in the blood-clotting cascade, are a family of Ca^{++}-dependent enzymes thought to be active in the covalent cross-linking of certain extracellular matrix proteins. They catalyze the linkage of specific glutamine and lysine side-chains by a transamidation reaction to form epsilon (gamma-glutamyl) lysine cross-links [112–114]. A role in bone matrix is suspected [115]. Methods of identification of transglutaminase-mediated cross-links rely on the use of tritiated putrescine or cadaverine to label candidate glutaminyl cross-linking sites. No new technique has been developed in the last decade. Several [^{3}H]putrescine-binding sites have been identified in collagens III, V, XI and XVI. The potential glutamine sites are present in the aminopropeptide of type III collagen, the non-triple-helical telopeptides of α1(V) and α1(XI) chains, and the N-terminal noncollagenous domain (NC11) of α1(XVI) chain [116–118]. However, none of the other cross-linking partners, the lysine sites, have been identified or suggested.

8.3 Tyrosine-Derived Cross-Links

The cuticles of *C. elegans* and other nemotodes consist of short-helix collagen polymers cross-linked by dityrosines and trityrosines [119, 120]. The lysyl oxidase mechanism does not operate. Instead a peroxidase is responsible. The enzyme catalyzing worm cuticle collagen cross-linking is a membrane-bound dual oxidase/peroxidase referred to as Duox [121]. A homologue to this enzyme is expressed in human tissues, but whether it has a similar function in generating extracellular di- and tri-tyrosine cross-links is unknown. Low levels of such cross-links can be found in vertebrate tissues but whether they are formed specifically or as an oxidative byproduct (e.g., in inflammation) is still unclear.

8.4 Types VIII and X Collagens

These homologous short chain molecules form hexagonal lattices. There is evidence that collagen type X, restricted to the hypertrophic and mineralized zones of growth plate cartilages, is cross-linked by the lysyl oxidase mechanism [122, 123], in addition to interchain disulfide bonding. However, site-specific cross-linked peptides need to be isolated to establish a role for lysyl oxidase-mediated cross-links. It is clear, nevertheless, that unusually strong hydrophobic interactions occur between the C-terminal globular domains in the hexagonal networks that these collagens form.

9
Outlook

The functional significance of the differences in cross-linking chemistry between tissue types is not clear. Imposed packing constraints on collagen molecules through the placement of cross-links may be more significant than the chemistry of the links themselves. On the other hand a reversible cross-linking chemistry may offer benefits, for example, in facilitating the mineralization of collagen fibrils in bone by allowing mineral crystallites to interdigitate and push apart the filamentous elements making up a fibril. Pyrrole cross-links are also thought to confer special qualities to bone collagen fibrils, perhaps related to the unique capacity of bone collagen to mineralize. The function of the pyrrole cross-links and significance of the link observed between the microscopic character of bone trabeculae and the ratio of pyridinoline to pyrrole cross-links need further study.

The functional significance in general of the trivalent cross-links, which can link three adjacent-collagen molecules, two in register and a third staggered by a 4D-overlap (where D=1/4.4 of the molecular length), is not clear. These permanent, end-products of hydroxylysine-aldehyde cross-linking are associated with tough connective tissues that bear high loads and are subject to minimal turnover. Whether they add mechanical strength of a particular quality, for example preventing lateral slippage between sub-elements of fibrils or offer a mechanism for interfibrillar bonding, is not known.

Perhaps the most important questions driving current research are aimed at understanding the cellular mechanisms that control tissue-specific differences in collagen cross-linking chemistry. Pivotal enzymes clearly include the three known lysyl hydroxylases, one of which (PLOD2) is believed to act on telopeptide lysines and so adapt the nascent collagen to cross-link via the hydroxylysine aldehyde pathway. This pathway is associated with normally tough connective tissues and with pathological fibrotic conditions. Molecular mechanisms that control the expression and activities of these regulatory gene products in different cell types will be important to define.

References

1. Exposito JY, Cluzel C, Garrone R, Lethias C (2002) Anat Rec 268:302
2. Boot-Handford RP, Tuckwell DS (2003) Bioessays 25:142
3. Bailey AJ, Paul RG, Knott L (1998) Mech Ageing Dev 106:1
4. Eyre DR, Paz MA, Gallop PM (1984) Annu Rev Biochem 53:717
5. Robins SP (1999) Fibrillogenesis and maturation of collagens. In: Seibel MJ, Robins SP, Bilezikian JP (eds) Dynamics of bone and cartilage metabolism. Academic Press, London, p 31
6. Eyre DR, Glimcher MJ (1971) Biochim Biophys Acta 243:525
7. Van Ness KP, Koob TJ, Eyre DR (1988) Comp Biochem Physiol B 91:531
8. Grandhee SK, Monnier VM (1991) Biol Chem 266:11649

9. Bailey AJ, Sims TJ, Avery NC, Miles CA (1993) Biochem J 296:489
10. Fu MX, Wells-Knecht KJ, Blackledge JA, Lyons TJ, Thorpe SR, Baynes JW (1994) Diabetes 43:676
11. Bailey AJ, Sims TJ, Avery NC, Halligan EP (1995) Biochem J 305:385
12. Sady C, Khosrof S, Nagaraj R (1995) Biochem Biophys Res Commun 214:793
13. Paul RG, Bailey AJ (1996) Int J Biochem Cell Biol 28:1297
14. Slatter DA, Paul RG, Murray M, Bailey AJ (1999) J Biol Chem 274:19,661
15. Slatter DA, Bolton CH, Bailey AJ (2000) Diabetologia 43:550
16. Piez KA (1968) Annu Rev Biochem 37:547
17. Tanzer ML (1973) Science 180:561
18. Tanzer ML, Waite JH (1982) Coll Relat Res 2:177
19. Bornstein P (2003) Matrix Biol 22:385
20. Hanson DA, Eyre DR (1996) J Biol Chem 27:26508
21. Wu JJ, Eyre DR (1984) Biochem 23:1850
22. Eyre DR, Apone S, Su JJ, Ericsson LH, Walsh KA (1987) FEBS Lett 220:337
23. Kuboki Y, Okuguchi M, Takita H, Kimura M, Tsuzaki M, Takakura A, Tsunazawa S, Sakiyama F, Hirano H (1993) Connect Tissue Res 29:99
24. Yamauchi M, London RE, Guenat C, Hashimoto F, Mechanic GL (1987) J Biol Chem 262:11,428
25. Yamauchi M, Chandler GS, Tanzawa H, Katz EP (1996) Biochem Biophys Res Commun 219:311
26. Mechanic GL, Katz EP, Henmi M, Noyes C, Yamauchi M (1987) Biochemistry 26: 3500
27. Henkel W, Rauterberg J, Stirtz T (1976) Eur J Biochem 69:223
28. Henkel W, Glanville R (1982) Eur J Biochem 122:205
29. Henkel W (1996) Biochem J 318:497
30. Wu JJ, Knigge PE, Eyre DR (1994) Trans Ortho Res Soc 19:131
31. Eyre DR, Wu JJ (1987) Type XI or 1α2α3α collagen. In: Mayne R, Burgeson RE (eds) Structure and function of collagen. Academic Press, Orlando, p 261
32. Niyibizi C, Eyre DR (1989) FEBS Lett 242:314
33. Mayne R, Brewton RG, Mayne PM, Baker JR (1993) J Biol Chem 268:9381
34. Wu JJ, Eyre DR (1995) J Biol Chem 270:18865
35. Niyibizi C, Eyre DR (1994) Eur J Biochem 224:943
36. Shaw LM, Olsen BR (1991) Trends Biochem Sci 16:191
37. Wu JJ, Woods PE, Eyre DR (1992) J Biol Chem 267:23007
38. Diab M, Wu JJ, Eyre DR (1996) Biochem J 314:327
39. Eyre DR, Wu JJ, Fernandes RJ, Pietka TA, Weis MA (2002) Biochem Soc Trans 30:844
40. Fernandes RJ, Schmid TM, Eyre DR (2003) Eur J Biochem 270:3243
41. Eyre DR, Pietka T, Weis MA, Wu JJ (2004) J Biol Chem 279:2568
42. Poole CA, Flint MH, Beaumont BW (1987) J Orthop Res 5:509
43. Mendler M, Eich-Bender SG, Vaughan L, Winterhalter KH, Bruckner P (1989) J Cell Biol 108:191
44. Vaughan-Thomas A, Young RD, Phillips AC, Duance VC (2001) J Biol Chem 276: 5303
45. Hagg R, Hedbom E, Mollers U, Aszodi A, Fassler R, Bruckner P (1997) J Biol Chem 272: 20,650
46. Fassler R, Schnegelsberg PN, Dausman J, Shinya T, Muragaki Y, McCarthy MT, Olsen BR, Jaenisch R (1994) Proc Natl Acad Sci USA 91:5070
47. Nakata K, Ono K, Miyazaki J, Olsen BR, Muragaki Y, Adachi E, Yamamura K, Kimura T (1993) Proc Natl Acad Sci USA 90:2870

48. Chapman KL, Briggs MD, Mortier GR (2003) Pediatr Pathol Mol Med 22:53
49. Fitzgerald J, Bateman JF (2001) FEBS Lett 505:275
50. Boute N, Exposito JY, Boury-Esnault N, Vacelet J, Noro N, Miyazaki K, Yoshizato K, Garrone R (1996) Biol Cell 88:37
51. Netzer KO, Suzuki K, Itoh Y, Hudson BG, Khalifah RG (1998) Protein Sci 7:1340
52. Fowler SJ, Jose S, Zhang X, Deutzmann R, Sarras MP Jr, Boot-Handford RP (2000) J Biol Chem 275:39,589
53. Ozbek S, Pertz O, Schwager M, Lustig A, Holstein T, Engel J (2002) J Biol Chem 277:49,200
54. Timpl R, Wiedemann H, van Delden V, Furthmayr H, Kuhn K (1981) Eur J Biochem 120:203
55. Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG (2003) N Engl J Med 348: 2543
56. Bailey AJ, Sims TJ, Light N (1984) Biochem J 218:713
57. Reddy GK, Hudson BG, Bailey AJ, Noelken ME (1993). Biochem Biophys Res Commun 190:277
58. Than ME, Henrich S, Huber R, Ries A, Mann K, Kuhn K, Timpl R, Bourenkov GP, Bartunik HD, Bode W (2002) Proc Natl Acad Sci USA 99:6607
59. Scott JE, Qian R, Henkel W, Glanville RW (1983) Biochem J 209:263
60. Bank RA, Robins SP, Wijmenga C, Breslau-Siderius LJ, Bardoel AF, van der Sluijs HA, Pruijs HE, te Koppele JM (1999) Proc Natl Acad Sci USA 96:1054
61. van der Slot AJ, Zuurmond AM, Bardoel AF, Wijmenga C, Pruijs HE, Sillence DO, Brinckmann J, Abraham DJ, Black CM, Verzijl N, DeGroot J, Hanemaaijer R, te Koppele JM, Huizinga TW, Bank RA (2003) J Biol Chem 278:40,967
62. Hautala T, Byers MG, Eddy RL, Shows TB, Kivirikko KI, Myllyla R (1992) Genomics 13:62
63. Szpirer C, Szpirer J, Riviere M, Vanvooren P, Valtavaara M, Myllyla R (1997) 8:707
64. Passoja K, Rautavuoma K, Ala-Kokko L, Kosonen T, Kivirikko KI (1998) Proc Natl Acad Sci USA 95:10,482
65. Ruotsalainen H, Sipila L, Kerkela E, Pospiech H, Myllyla R (1999) Matrix Biol 18:325
66. Yeowell HN, Walker L (1999) Matrix Biol 18:179
67. Hyland J, Ala-Kokko L, Royce P, Steinmann B, Kivirikko KI, Myllyla R (1992) Nat Genet 2:228
68. Yeowell HN, Walker LC (1997) Proc Assoc Am Phys 109:383
69. Yeowell HN, Walker LC (2000) Mol Genet Metab 71:212
70. Steinmann B, Eyre DR, Shao P (1995) Am J Hum Genet 57:1505
71. Eyre D, Shao P, Weis MA, Steinmann B (2002) Mol Genet Metab 76:211
72. Heikkinen J, Risteli M, Wang C, Latvala J, Rossi M, Valtavaara M, Myllyla R (2000) J Biol Chem 275:36,158
73. Wang C, Luosujarvi H, Heikkinen J, Risteli M, Uitto L, Myllyla R (2002) Matrix Biol 21:559
74. Wang C, Risteli M, Heikkinen J, Hussa AK, Uitto L, Myllyla R (2002) J Biol Chem 277: 18,568
75. Uzawa K, Grzesik WJ, Nishiura T, Kuznetsov SA, Robey PG, Brenner DA, Yamauchi M (1999) J Bone Miner Res 14:1272
76. Mercer DK, Nicol PF, Kimbembe C, Robins SP (2003) Biochem Biophys Res Commun 307:803
77. Knott L, Bailey AJ (1998) Bone 22:181
78. Yamauchi M, Katz EP (1993) Connect Tissue Res 29:81
79. Banse X, Devogelaer JP, Lafosse A, Sims TJ, Grynpas M, Bailey AJ (2002) Bone 31:70
80. Knott L, Tarlton JF, Bailey AJ (1997) Biochem J 322:535

81. Wassen MH, Lammens J, te Koppele JM, Sakkers RJ, Liu Z, Verbout AJ, Bank RA (2000) J Bone Miner Res 15:1776
82. Paschalis EP, Verdelis K, Doty SB, Boskey AL, Mendelsohn R, Yamauchi M (2001) J Bone Miner Res 16:1821
83. Paschalis EP, Recker R, DiCarlo E, Doty SB, Atti E, Boskey AL (2003) J Bone Miner Res 18:1942
84. Blank RD, Baldini TH, Kaufman M, Bailey S, Gupta R, Yershov Y, Boskey AL, Coppersmith SN, Demant P, Paschalis EP (2003) Connect Tissue Res 44:134
85. Brady JD, Robins SP (2001) J Biol Chem 276:18812
86. Adamczyk M, Johnson DD, Reddy RE (2000) Bioconjug Chem 11:124
87. Kleter GA, Damen JJ, Kettenes-van den Bosch JJ, Bank RA, te Koppele JM, Veraart JR, ten Cate JM (1998) Biochim Biophys Acta 1381:179
88. Eriksen HA, Sharp CA, Robins SP, Sassi ML, Risteli L, Risteli J (2004) Bone 34:720
89. Bank RA, te Koppele JM, Janus GJ, Wassen MH, Pruijs HE, Van der Sluijs HA, Sakkers RJ (2000) J Bone Miner Res 2000 15:1330
90. Delmas PD (1993) J Bone Miner Res 8:S549
91. Eyre DR (1995) Acta Orthop Scand Suppl 266:266
92. Christenson RH (1997) Clin Biochem 30:573
93. Garnero P, Delmas PD (1998) Endocrinol Metab Clin North Am 27:303
94. Watts NB (1999) Clin Chem 45:1359
95. Beardsworth LJ, Eyre DR, Dickson IR (1990) J Bone Min Res 5:671
96. Uebelhart D, Gineyts E, Chapuy MC, Delmas PD (1990) Bone Miner 8:87
97. Robins SP, Black D, Paterson CR, Reid DM, Duncan A, Seibel MJ (1991) Eur J Clin Invest 21:310
98. Seibel MJ, Robins SP, Bilezikian JP (1992) Trends Endocrinol Metab 3:263
99. Eastell R, Colwell A, Hampton L, Reeve J (1997) J Bone Miner Res 12:59
100. Hanson DA, Weis MA, Bollen AM, Maslan SL, Singer FR, Eyre DR (1992) J Bone Miner Res 7:1251
101. Risteli J, Elomaa I, Niemi S, Novamo A, Risteli L (1993) Clin Chem 39:635
102. Bollen AM, Eyre DR (1994) Bone 15:31
103. Eyre DR, Atley LM, Wu JJ (2002) Collagen cross-links as markers of bone and cartilage degradation In: Hascall VC, Kuettner KE (eds) The many faces of osteoarthritis. Birkhäuser Verlag, Basel, Switzerland, p 275
104. Clemens JD, Herrick MV, Singer FR, Eyre DR (1997) Clin Chem 43:2058
105. Atley LM, Mort JS, Lalumiere M, Eyre DR (2000) Bone 26:241
106. Eyre DR (1992) US Patent 5,140,103
107. Lohmander LS, Atley LM, Pietka TA, Eyre DR (2003) Arthritis Rheum 48:3130
108. Eyre DR, Atley LM, Vosberg-Smith K, Shaffer K, Ochs V, Clemens D (2000) Biochemical markers of bone and cartilage degradation. In: Goldberg M, Boskey A, Robinson C (eds) Chemistry and biology of mineralized tissues. AAOS, p 347
109. Atley LM, Sharma L, Clemens JD, Shaffer K, Pietka TA, Riggins JA, Eyre DR (2000) Trans Orthop Res Soc 25:168
110. O'Kane JW, Atley L, Hawley C, Teitz C, Eyre DR (2001) Orthop Res Rep, UW, Dept of Orthopedics and Sports Medicine, 2001:3
111. Wu JJ, Eyre DR, Slayter HS (1987) Biochem J 248:373
112. Folk JE (1980) Ann Rev Biochem 49:517
113. Griffin M, Casadio R, Bergamini C (2002) Biochem J 368:377
114. Lorand L, Graham RM (2003) Nat Rev Mol Cell Biol 4:140
115. Aeschlimann D, Mosher D, Paulsson M (1996) Semin Thromb Hemost 22:437
116. Bowness JM, Folk JE, Timpl R (1987) J Biol Chem 262:1022

117. Kleman JP, Aeschlimann D, Paulsson M, van der Rest M (1995) Biochemistry 34:13,768
118. Akagi A, Tajima S, Ishibashi A, Matsubara Y, Takehana M, Kobayashi S, Yamaguchi N (2002) J Invest Dermatol 118:267
119. Fetterer RH, Rhoads ML, Urban JF Jr (1993) J Parasitol 79:160
120. Yang J, Kramer JM (1999) J Biol Chem 274:32,744
121. Edens WA, Sharling L, Cheng G, Shapira R, Kinkade JM, Lee T, Edens HA, Tang X, Sullards C, Flaherty DB, Benian GM, Lambeth JD (2001) J Cell Biol 154:879
122. Rucklidge GJ, Milne G, Robins SP (1996) Matrix Biol 15:73
123. Orth MW, Luchene LJ, Schmid TM (1996) Biochem Biophys Res Commun 219:301